



44/15

Cranfield Institute of Technology  
Centre for Transport Studies

Ph.D Thesis

Petros C. Vythoulkas

Modelling Dynamic Stochastic User Equilibrium  
for Urban Road Networks

Supervisor

Ian G. Black

January 1991

This thesis is submitted for the degree of Doctor of Philosophy

στους γονείς μου  
κωνσταντίνο και λυγερή

## ABSTRACT

In this study a dynamic assignment model is developed which estimates travellers' route and departure time choices and the resulting time varying traffic patterns during the morning peak. The distinctive feature of the model is that it does not restrict the geometry of the network to specific forms.

The proposed framework of analysis consists of a travel time model, a demand model and a demand adjustment mechanism. Two travel time models are proposed. The first is based on elementary relationships from traffic flow theory and provides the framework for a macroscopic simulation model which calculates the time varying flow patterns and link travel times given the time dependent departure rate distributions; the second is based on queueing theory and models roads as bottlenecks through which traffic flow is either uncongested or fixed at a capacity independent of traffic density. The demand model is based on the utility maximisation decision rule and defines the time dependent departure rates associated with each reasonable route connecting the O-D pairs of the network, given the total utility associated with each combination of departure time and route. Travellers' choices are assumed to result from the trade-off between travel time and schedule delay and each individual is assumed to first choose a departure time  $t$ , and then select a reasonable route, conditional on the choice of  $t$ . The demand model has therefore the form of a nested logit. The demand adjustment mechanism is derived from a Markovian model, and describes the day-to-day evolution of the departure rate distributions. Travellers are assumed to modify their trip choice decisions based on the information they acquire from recent trips. The demand adjustment mechanism is used in order to find the equilibrium state of the system, defined as the state at which travellers believe that they cannot increase their utility of travel by unilaterally changing route or departure time.

The model outputs exhibit the characteristics of real world traffic patterns observed during the peak, i.e., time varying flow patterns and travel times which result from time varying departure rates from the origins. It is shown that increasing the work start time flexibility results in a spread of the departure rate distributions over a longer period and therefore reduces the level of congestion in the network. Furthermore, it was shown that increasing the total demand using the road network results in higher levels of congestion and that travellers tend to depart earlier in an attempt to compensate for the increase in travel times. Moreover, experiments using the queueing theory based travel time model have shown that increasing the capacity of a bottleneck may cause congestion to develop downstream, which in turn may result in an increase of the average travel time for certain O-D pairs. The dynamic assignment model is also applied to estimate the effects that different road pricing policies may have on trip choices and the level of congestion; the model is used to demonstrate the development of the shifting peak phenomenon. Furthermore, the effect of information availability on the traffic patterns is investigated through a number of experiments using the developed dynamic assignment model and assuming that guided drivers form a class of users characterised by lower variability of preferences with respect to route choice.

## **ACKNOWLEDGEMENTS**

I would like to express my gratitude and sincere thanks to :

All the members of staff of the Centre for Transport Studies, Cranfield Institute of Technology, and all my friends at Cranfield for their support.

My supervisor Ian G. Black who encouraged me to start this study and whose constructive criticism, attention and interest have contributed most to the development of this work.

All the members of staff of the Transport Studies Unit, University of Oxford, for their help during the final but long stage of this study.

Professor Moshe Ben Akiva from Massachusetts Institute of Technology, and Professor Andre de Palma from Northwestern University whose comments and suggestions have been very valuable and constructive.

My friend Andreas for his continuous support since "the high school years" and for his invaluable help which enabled me to meet the deadlines.

Finally my thanks go to my parents and my sister whose constant encouragement, support and advice were the most important during all the stages of my education.

## Contents

Abstract . . . . .	iii
Acknowledgements . . . . .	iv
<b>1. Introduction.</b>	
1.1 Urban traffic congestion and transportation network analysis . . . . .	2
1.2 Static equilibrium assignment: Basic concepts and criticisms.. . . . .	3
1.3 Dynamic assignment: Basic concepts and rationale for further developments. . . . .	7
1.4 Setting the objectives. . . . .	10
1.5 Structure of the study . . . . .	11
<b>2. Multiple Choice Models.</b>	
2.1 Introduction . . . . .	15
2.2 The utility maximisation concept. . . . .	17
2.3 The multinomial probit . . . . .	20
2.4 The multinomial logit. . . . .	23
2.5 The continuous logit . . . . .	27
2.6 The nested logit . . . . .	28
2.7 Dynamic models of choice . . . . .	34
2.8 Satisfaction and the expected disutility of travel . . . . .	37
2.9 Summary . . . . .	40
<b>3. Static Assignment Models.</b>	
3.1 Introduction . . . . .	44
3.2 Network representation . . . . .	48
3.3 Link performance functions . . . . .	50
3.4 Heuristic equilibration techniques . . . . .	53
<i>capacity restraint</i> . . . . .	54
<i>incremental assignment</i> . . . . .	55
3.5 User equilibrium assignment. . . . .	56
<i>program formulation</i> . . . . .	56
<i>equivalency conditions</i> . . . . .	56
<i>uniqueness conditions</i> . . . . .	58
<i>solution algorithm</i> . . . . .	58
<i>extension of user equilibrium</i> . . . . .	60

3.6	Stochastic Equilibrium . . . . .	61
3.6.1	Stochastic network loading . . . . .	61
3.6.2	Stochastic user equilibrium assignment. . . . .	64
	<i>program formulation</i> . . . . .	64
	<i>equivalency conditions</i> . . . . .	65
	<i>uniqueness conditions</i> . . . . .	66
	<i>the solution algorithm</i> . . . . .	66
	<i>extensions of stochastic user equilibrium</i> . . . . .	67
3.7	Summary . . . . .	68
<b>4.</b>	<b>Peak Period Work Trip Scheduling.</b>	
4.1	Introduction . . . . .	72
4.2	The peaking nature of traffic flows and travel times . . . . .	74
	<i>traffic flow variability</i> . . . . .	75
	<i>travel time variability</i> . . . . .	77
4.3	Work trip scheduling . . . . .	79
4.4	Managing the work trip scheduling . . . . .	87
4.5	Summary . . . . .	92
<b>5.</b>	<b>A Review of Dynamic Network Analysis.</b>	
5.1	Introduction . . . . .	96
5.2	Travel time models . . . . .	100
5.2.1	Single O-D pair networks. . . . .	100
	<i>the deterministic queueing theory approach</i> . . . . .	100
	<i>the traffic flow theoretic approach.</i> . . . . .	102
	<i>the Markov chain approach</i> . . . . .	103
5.2.2	Urban transportation corridors . . . . .	105
	<i>traffic flow simulator</i> . . . . .	105
	<i>the traffic flow models</i> . . . . .	106
5.3	The Interaction between Demand and System's Performance . . . . .	107
5.3.1	The Markov chain formulation . . . . .	108
5.3.2	Utility maximisation approach . . . . .	110
	<i>deterministic equilibrium</i> . . . . .	110
	<i>stochastic equilibrium</i> . . . . .	113
5.3.3	Boundedly rational user equilibrium . . . . .	117
5.4	Empirical research . . . . .	120
5.5	Summary . . . . .	129

<b>6. A Dynamic Stochastic Assignment Model for General Networks.</b>	
6.1 Introduction . . . . .	134
6.2 The dynamic route choice set . . . . .	137
6.2.1 The set of reasonable paths . . . . .	138
6.2.1.1 <i>Dynamic shortest path algorithm.</i> . . . . .	139
6.3 The travel time model . . . . .	144
6.3.1 The traffic flow model . . . . .	145
6.3.2 The speed-density relationships. . . . .	149
6.3.3 Formulation of the time dependent O-D travel times. . . . .	151
6.3.4 The queueing theory approach . . . . .	153
6.4 The demand model. . . . .	156
6.4.1 Specification of the utility function . . . . .	159
<i>disutility of travel time.</i> . . . . .	159
<i>disutility of schedule delay.</i> . . . . .	159
<i>total utility of travel</i> . . . . .	161
6.4.2 The time dependent departure rates. . . . .	161
6.5 The demand adjustment mechanism . . . . .	162
6.6 The framework of the simulation model : . . . . .	172
6.7 Summary . . . . .	176
<b>7. Formulating dynamic assignment as a mathematical program.</b>	
7.1 Introduction . . . . .	181
7.2 The DSUE program formulation . . . . .	185
7.3 Equivalence conditions . . . . .	186
7.4 Uniqueness conditions . . . . .	190
7.5 Solution algorithm . . . . .	199
7.5.1 The method of successive averages . . . . .	202
7.6 Summary . . . . .	205
<b>8. Experimental analysis of the DSUE model.</b>	
8.1 Introduction . . . . .	208
8.2 The base case experiments using the traffic flow model . . . . .	225
8.3 Sensitivity to different work start time flexibilities . . . . .	225
8.4 Sensitivity to different levels of variability of preferences. . . . .	232
8.5 Sensitivity to different levels of demand. . . . .	238

**PAGE  
NUMBERING  
AS ORIGINAL**



8.6	Simulation experiments using the queueing theory model . . . . .	245
8.7	Comparison of the demand adjustment and the equivalent formulation approach . . . . .	250
8.8	Summary . . . . .	252
<b>9.</b>	<b>Using the DSUE model for policy analysis</b>	
9.1	Introduction. . . . .	255
9.2	Evaluating the effects of road pricing. . . . .	256
9.3	Forecasting the effects of drivers information systems. . . . .	272
9.4	Summary . . . . .	279
<b>10.</b>	<b>Conclusions</b>	
10.1	Summary and Conclusions . . . . .	282
10.2	Limitations and further research . . . . .	287
<b>References</b>	. . . . .	<b>291</b>
<b>Appendix 1</b>	. . . . .	<b>303</b>
<b>Appendix 2</b>	. . . . .	<b>304</b>

# 1 introduction

### **1.1 Urban traffic congestion and transportation network analysis.**

The separation between residential and employment areas has created an important characteristic of the urban transportation activities. This is the considerable movement of commuters (initially widely distributed throughout the urban area) who have to travel to and from the Central Business District (CBD), where the majority of workplaces are located. Furthermore, the practice of most working hours being similar through the whole spectrum of jobs leads to large peaks in the daily profile of transport demand, since people travel between home and work at approximately the same times during the day. This heavy concentration of trips both in terms of time and space gives rise to traffic congestion with its inherent problems and inconveniences.

The dramatic increase of urban traffic congestion is not any more a characteristic of major urban areas. Medium sized cities or even smaller urban areas experience levels of congestion which cost a significant amount of resources in terms of lost time from delays, contribute to serious environmental problems, and subsequently result in a decline of the quality of living and a restraint of economic growth in urban areas.

Traffic engineering techniques designed to reduce the adverse impacts of urban traffic congestion can be classified into three general categories (Rathi and Lieberman, 1989):

- Measures designed to *expand the infrastructure* aimed at providing increased capacity of the road networks, either by building additional facilities or by physically altering the existing ones.
- *Demand management* measures designed to reduce or to spread demand over less congested periods. Policies that are often adopted include traffic restrictions, road pricing schemes, flexible and staggered working hours, and others.
- Measures designed to *maximise the utilisation* of the available capacity by either minimising the capacity reducing factors, (e.g. parking, standing and stopping

control) or maximising the utilisation of the existing infrastructure (e.g. advanced signal control techniques, motorist information systems).

It is widely believed that implementation of one of the above measures alone does not provide a solution to the urban traffic problem, and strategies consisting of a combination of these measures should be therefore applied. However implementation of such strategies may require a significant amount of resources, or may involve alterations of the social and economic activities which are difficult to achieve.

Thus, the adversity of the urban traffic congestion problem on the one hand, and the high amount of resources required to reduce its impacts on the other, argue for the development of procedures which can accurately estimate and analyse the effects of the different strategies used to combat congestion. Within this context and over the last decades, research on the area of transportation network analysis has made significant advances in developing traffic assignment models aiming at providing an accurate representation of the complex interrelationships between drivers' behaviour and transportation networks performance. Traffic assignment procedures are primarily concerned with the analysis of road networks; they deal mainly with the allocation of an origin-destination matrix onto alternative routes of a network and estimate the traffic flow levels, delays and travel times in the links formulating that network.

Traffic assignment models fall in two general categories depending on the way they view the time dimension of travel demand:

- *Static* models which analyse the assignment problem within the space domain and ignore the time dimension of travel demand, and
- *Dynamic* models which take into account the time variability of travel demand.

The basic concepts of static and dynamic assignment modelling will be briefly discussed in the two following sections.

## **1.2 Static equilibrium assignment: Basic concepts and criticisms**

This section will introduce the basic concepts of the static approach to traffic assignment modelling and will discuss some criticisms against it. This, in order to support the arguments for the need of developing the area of research concerned with dynamic network analysis. A review of the research on static assignment modelling will be presented in chapter three.

As stated earlier, traffic assignment procedures deal with the allocation of an O-D matrix onto alternative routes of the network that is analysed. To find the solution to this problem it is necessary to define the rule by which travellers choose a route. It is reasonable to assume that every driver will try to minimise his travel time when travelling from origin to destination. Given the demand between an O-D pair, the question is how the travellers will be distributed among the possible paths connecting this O-D pair. If all of them were to take the shortest path (in terms of travel time), congestion would develop on it. Consequently travel time on this path might increase up to a point where it is no longer the minimum travel time path, and therefore some of the travellers would divert to another path. This alternative path can however be also congested, and so on. The interaction of congestion and travel decisions will finally result in an equilibrium flow pattern. Since congestion increases with flow and trips are discouraged by congestion, this interaction can be modelled as a process of reaching an equilibrium between congestion and travel decisions. The concept of equilibrium, in the transportation planning context, was first† introduced, by Wardrop (1952) who stated that the traffic on a network distributes itself in such a way that *"the journey times in all the routes actually used are equal, and less than those which would be experienced by a single vehicle on any unused route . . . since it might be assumed that traffic will tend to settle down into an equilibrium situation in which no traveller can reduce his travel time by choosing a new route"*. This condition is known as the *user equilibrium (UE) condition* or *Wardrop's first principle*.

The user-equilibrium condition includes several assumptions which can be criticised as not realistic. For example travellers are assumed to have full information on the travel times on every route, and to make the correct decisions. However in reality, every traveller may *perceive* a different travel time over the same link, and thus a more realistic decision rule should be that each traveller tries to minimise his perceived travel time. This assumption leads to the *stochastic user equilibrium (SUE) condition* which was introduced by Daganzo and Sheffi (1977) who stated that *"In a stochastic user equilibrium network no user believes he can improve his travel time by unilaterally changing routes"*. The stochastic user equilibrium is a generalisation of the user equilibrium definition, since the perceived travel time can be formulated as a random variable distributed across the population of drivers, with mean equal to the actual

---

† Beckman *et al.* (1956) points out that the principle of traffic equilibrium was demonstrated by Knight F.H. (1924) in the article: 'Some Fallacies in the Interpretation of Social Cost'. *Quarterly Journal of Economics*, Vol 38 pp. 582-606. He supported his argument using a network consisting of two roads; one with high capacity and bad geometric characteristics and one with low capacity and good geometric characteristics.

travel time. Thus if the perception of travel time is accurate, then the SUE flow pattern will be identical to the deterministic UE pattern.

The static equilibrium assignment models estimate the morning and evening rush hour flows by assigning a fraction of the daily demand on the network. Peak periods are arbitrarily defined to be fixed time intervals within which demand is assumed to be uniformly distributed, or in other words that the trip rates between the O-D pairs are constant during the period of analysis. Furthermore, the static approach implies that the traffic flow patterns are time invariable and therefore the link flows and link travel times are constant during the modelled time interval.

Thus, a concise description of the distinctive features of static assignment can be given with the following statement:

" . . . a standard static network equilibrium model represents the flow patterns, during a *fixed time interval* - a peak period, or a peak hour. The temporal, or time of day, distribution of the traffic is assumed to be fixed and during the modelled time interval the traffic is assumed to be *uniformly distributed* . . . ". Ben-Akiva (1985).

However, as will be demonstrated in chapter four, traffic volumes vary in both space and time. Furthermore, under the assumptions inherent in static assignment procedures, travellers respond to congestion solely by choosing the route that minimise their own travel time. Yet, observations support the view that travel times are highly dependent on the time of the day at which individuals decide to travel. Thus, conventional network analysis does not take into account another dimension of choice also available to travellers for combating congestion; this is the decision on departure time.

Static network analysis was criticised by several researchers :

". . traffic assignments cannot indicate the locations and extents of queues or the delays associated with them. Because queueing can be of major importance in peak period expressway operations in downtown areas, the assignments can be grossly inaccurate in predicting peak period operating speeds . ." (Lisco, 1983)

". . standard static network equilibrium formulation fails to capture essential features of traffic congestion...highway travel times are convex function functions of the traffic flow; therefore by Jensen's inequality† a *static traffic assignment systematically underestimates travel times* . . ." (Ben-Akiva, 1985)

---

† The travel time functions, used in traffic assignment, will be discussed in section 3.3. Jensen's inequality (Rao, 1973) is expressed as :

$$E[f(x)] \geq f[E(x)]$$

where  $x$  is a random variable following a given probability distribution, and  $f(x)$  is a convex function.

" . . no consideration of departure time or the dynamic nature of transportation system performance during the peak period is included. The peak period of travel is typically described by a set of uniform travel characteristics, representing an unrealistic but analytically convenient compromise . . ." (Hendrickson and Plank, 1984)

" . . In traditional assignment models . . . Cars between each O-D pair are assigned to the links which belong to certain route. As these links do not have a time dimension, the implicit assumption is made that cars are presented on all links at the same time. So cars which in reality are held in a certain bottleneck can in the calculation also cause a congestion downstream . ." (Hammerslag, 1988)

" . . Clearly, a uniform peak period travel pattern is a fictitious concept . . . It is evident that the temporal characteristics of traffic congestion which play an important role in the determination of operating speeds are not captured by steady state network models . . ." (Ben-Akiva and de Palma, 1987)

Static traffic assignment models are mainly used to evaluate alternative strategies to relief traffic congestion. However the assumption of the arbitrarily defined fixed peak period, seriously limits the validity of their predictions. Thus for example, by the usual method of analysis, adding capacity to a congested transportation facility would result in a decrease in travel times within the fixed peak period, by an amount depending on the additional capacity provided, and would possibly produce free-flow conditions. Empirical observations, though, support a quite different view. Small (1982) comments that " . . . traffic counts on the San Fransisco-Oakland Bay Bridge, before and after the opening of a parallel rapid transit line in 1974, showed that this considerable expansion of total capacity resulted in a *negligible change in the level of congestion* on the bridge, but a *substantial decrease in its duration* . . .".

Kroes *et al.* (1987) also reports that expansion of capacity may uncover *latent demand* for travel during the peak period. To support his argument he gives the example of Lek Bridge at Viannen, Netherlands, where a significant component of this latent demand was due to travellers who had adjusted their departure time in order to avoid peak-hour congestion, but returned to the peak after the capacity expansion.

It is therefore evident that investments involving high costs do not necessarily result in any significant decrease in the level of congestion, since travellers shift their departure times. They obviously do so in order to reduce their *schedule delay*, defined as the difference between the desired and the actual arrival time at the destination. Thus, despite the fact that capacity expansion may not result in the expected reduction in travel times, it certainly contributes to a better level of service due to the lower levels of schedule delay. Cost-benefit analyses based on static assignment models, will then

contain inaccuracies attributed not only to the erroneous predictions of travel time, but also to the omission of the benefits derived from the altered trip schedules.

### **1.3 Dynamic Assignment: Basic concepts and rationale for further developments.**

The evidence that departure time choice decisions directly influence peak period patterns and congestion levels in transportation facilities has resulted in an increasing interest in modelling trip timing decisions. Over the last decade, several researchers have analysed the departure time choice problem by developing econometric demand models of work trip scheduling, which will be discussed in chapter four. All these models, were calibrated with real data, and estimate the probability of selecting each of the alternative departure times. These efforts constitute a major contribution to the better understanding of the trip timing decision problem, since they provide estimates of the trade-off between schedule delay and travel time; however they assume that the characteristics of the transportation system are given and known to every user, rather than dependent on his own as well as the other users' decisions. In other words, they consider only the demand side of the problem, do not treat the effect of travel demand on travel times, and therefore omit the interaction between system's performance and users' decisions.

Furthermore, much research has been devoted to understanding travellers' departure time decisions in studies concerned with reducing peak-period congestion through demand side measures such as flexible and staggered work hours and road pricing strategies. An understanding of users' departure time and route selection decisions and their interrelation with congestion, is of great importance. The development of assignment models to incorporate these interrelations, and to have appropriate predictive capability is therefore also necessary for the design and evaluation of alternative measures for coping with peak-period congestion.

Moreover, the development of such models becomes a necessity in the light of the current interest in the potential of route guidance systems. Up to date the majority of the research effort on the design of driver information systems has concentrated on the "hardware" side of the problem. However, a crucial component, which also needs a thorough examination, is the traffic model in the central control centre, which will calculate the optimum routes for current and expected network conditions. This model will require a trip assignment procedure, which can advise routes in space and time, based on expected time varying traffic conditions.



The major shortcomings of static assignment models, i.e. their failure to represent the time varying characteristics of traffic flow and travel demand and therefore to provide an accurate picture of the traffic conditions during the peak period congestion, has led to the increasing interest in modelling the time-varying nature of travel demand during the peak period. The developed models were labelled by several authors as *dynamic network models*, and can be classified in two broad categories :

- The first refers to the models which estimate the time-varying flow patterns and travel times, given a dynamic (time dependent) O-D matrix.
- the second refers to models which require only the total O-D trip matrix, and the passengers' desired arrival times at the destinations, and predict the temporal distribution of the demand, i.e. the time dependent distribution of departure rates for each O-D pair, and then estimate the time varying traffic flow patterns.

Models included in the first category are seriously limited by their input data requirements; a time dependent O-D matrix is difficult to obtain and even more difficult to forecast. Furthermore, these models cannot simulate the effects of variable work schedules on the temporal distribution of demand, and therefore on the time-varying flows and travel times in the network. They cannot also predict the altering of work trip scheduling and therefore the shift of off-peak demand into the peak period, after expanding the capacity of existing transportation facilities.

The models included in the second category are not subject to the above limitations and constitute the main topic of interest within this study. To find the solution to the problem they address, it is necessary to define the rule by which travellers select which time to start their journey and which route to follow. Decision rules and choice models will be discussed in chapter two, and their specific formulation required for the route and departure time decision problem is derived from the material that will be presented in chapters four and five. Here a brief description of the main assumption used in the development of these models is given:

- travellers make their decisions based on utility maximisation decision rules,
- they have the choice between an on time arrival with a long travel time and a late or an early arrival with a shorter travel time,
- their choices are assumed to result from the trade-off between travel time and schedule delay, and thus
- the disutility they experience is attributable to travel time and schedule delay.

Given the demand between an O-D pair, the question is then how the travellers will be distributed between alternative departure times and among the possible paths connecting this O-D pair. If all of them were to take the shortest path (in terms of travel time), and depart at such time so that they arrive on time at their destination, congestion would develop on it. Consequently travel time on this path and for this particular departure time might increase up to a point where it is no longer the minimum travel time path and it does not imply on time arrivals; therefore some of the travellers would divert to another departure time and/or path. This alternative path can however be also congested at the time that the traveller follows it, and so on. The interaction of congestion and travel decisions will finally result in an equilibrium flow pattern.

Thus, in an analogy to the UE condition used in static assignment, a dynamic network equilibrium (DUE), can be expressed as an extension of Wardrop's first principle to the case of route and departure time choice. It is defined as the equilibrium state at which no driver can reduce his disutility of travel by selecting a different route or departure time. Moreover, to relax the restrictive assumptions (e.g. full information on travel times on each route and for each departure time and, travellers' ability to make correct decisions) inherent in the deterministic equilibrium hypothesis, stochastic formulations can be used. Thus, Dynamic Stochastic User Equilibrium (DSUE), can be then regarded as a combination of DUE and SUE, and is defined as the state at which travellers believe that they cannot any more reduce their disutility of travel by unilaterally changing route or departure time.

Research on the area of dynamic stochastic user equilibrium modelling is still at its early stages of development. In the review of the research on DSUE assignment models that will be presented in chapter five, it is shown that existing dynamic assignment models can handle only simple network forms and restrict the trip decisions to departure time choice; no route choice actually exist, with the exception of the model developed by Ben-Akiva *et al.* (1986a,b) which however can handle only an one O-D pair network connected by parallel routes.

The need to develop dynamic assignment models which can analyse general networks was pointed out by several researchers. Amongst others, Friesz (1985) suggests that "*..among improvements to steady-state network equilibrium models most likely to enhance their predictive capability are the needs to : include dynamic considerations....*", Ben-Akiva (1985) and Ben-Akiva and de Palma (1987) argue for the need "*..to establish the feasibility of applying dynamic equilibrium models to more realistic networks..*"

Furthermore, dynamic network analysis has been the main topic in a number of conferences, e.g. the USA-Italy Joint Seminar on "Urban Traffic Networks: Dynamic Control and Flow Equilibrium", and was considered as one of the major research tasks set by the DRIVE programme of the Commission of European Communities.

#### 1.4 Setting the objectives

Having discussed the concept and the current state of the research in dynamic stochastic network equilibrium analysis, the objectives of this work are presented below.

The main objective of this thesis is to:

*Develop a Dynamic Stochastic User Equilibrium assignment model which can analyse general network forms. This model will predict the commuters' route and departure time choices and will estimate the time varying traffic flow patterns during the morning peak period. The distinctive feature of the model is that it does not restrict the geometry of the network to certain forms.*

In order to facilitate the derivation of the DSUE assignment model the following secondary objectives are defined:

- to review the existing research on multiple choice analysis and dynamic and static assignment.
- to describe the peak period phenomena and travellers' behaviour with respect to route and departure time choice decisions.

To demonstrate the significance of this work, several simulation experiments are conducted with the aim to:

- present the outputs of the model and compare them with the outputs from static formulations
- show how the DSUE model can be used to evaluate the effects of a road pricing scheme and a route guidance system, on the traffic patterns during the morning peak period.

## 1.5 Structure of the study

This section presents the development of the chapters of this thesis and their dependencies. The chapter dependency is also illustrated in figure 1.1 where the emphasised arrows illustrate the stronger relationships.

Since a major component of the DSUE assignment problem deals with the estimation of drivers' choice decisions, it is necessary to describe the concepts and the currently available basic models used in the analysis of multiple choice behaviour. This is done in chapter two with the emphasis given to the utility maximisation approach, and particularly to the development and properties of the logit formulations. The material provided in this chapter is required to comprehend the development of the modelling procedures described in subsequent chapters. The continuous, nested and dynamic logit formulations presented in sections 2.5 2.6 and 2.7 respectively will be used in the development of the demand models and the demand adjustment mechanism presented in chapters 5 and 6. Furthermore the concepts of satisfaction and the expected disutility of travel (section 2.8) will be used in the equivalent program formulation of the static and dynamic stochastic assignment problem described in section 3.6 and chapter 7 respectively.

Chapter 3 provides the necessary background information needed for the analysis of the traffic assignment model and provides a review of the research on static traffic assignment. The focus of the chapter is on static equilibrium assignment models both deterministic and stochastic. The review is neither historical nor comprehensive but is required for the understanding of the development of the equivalent program formulation of the static stochastic assignment model given in section 3.7. The material presented in this section is required to comprehend the formulation of DSUE as an equivalent optimisation program and to prove the uniqueness of its solution as they are described in chapter 7.

Chapter 4 addresses the topic of peak period work trip scheduling. It demonstrates the characteristics of demand and travel time peaking in transportation systems and reviews some studies on work trip scheduling which are required to formulate the utility functions used in the choice models developed in chapters 5 and 6. The chapter concludes with a presentation of different strategies which are used to alter work trip scheduling, in order to reduce the effects of peaks by spreading the demand over a longer period.

Chapter 5 reviews the currently developed dynamic network analysis models. These models take into account the spatial and time-of-day variability of network congestion but can only analyse specific network forms. The chapter deals with both deterministic and stochastic dynamic assignment models and discusses the utility maximisation as well as the bounded rationality user equilibrium approach. The review is rather comprehensive and provides a framework for developing the DSUE assignment model presented in chapter 6.

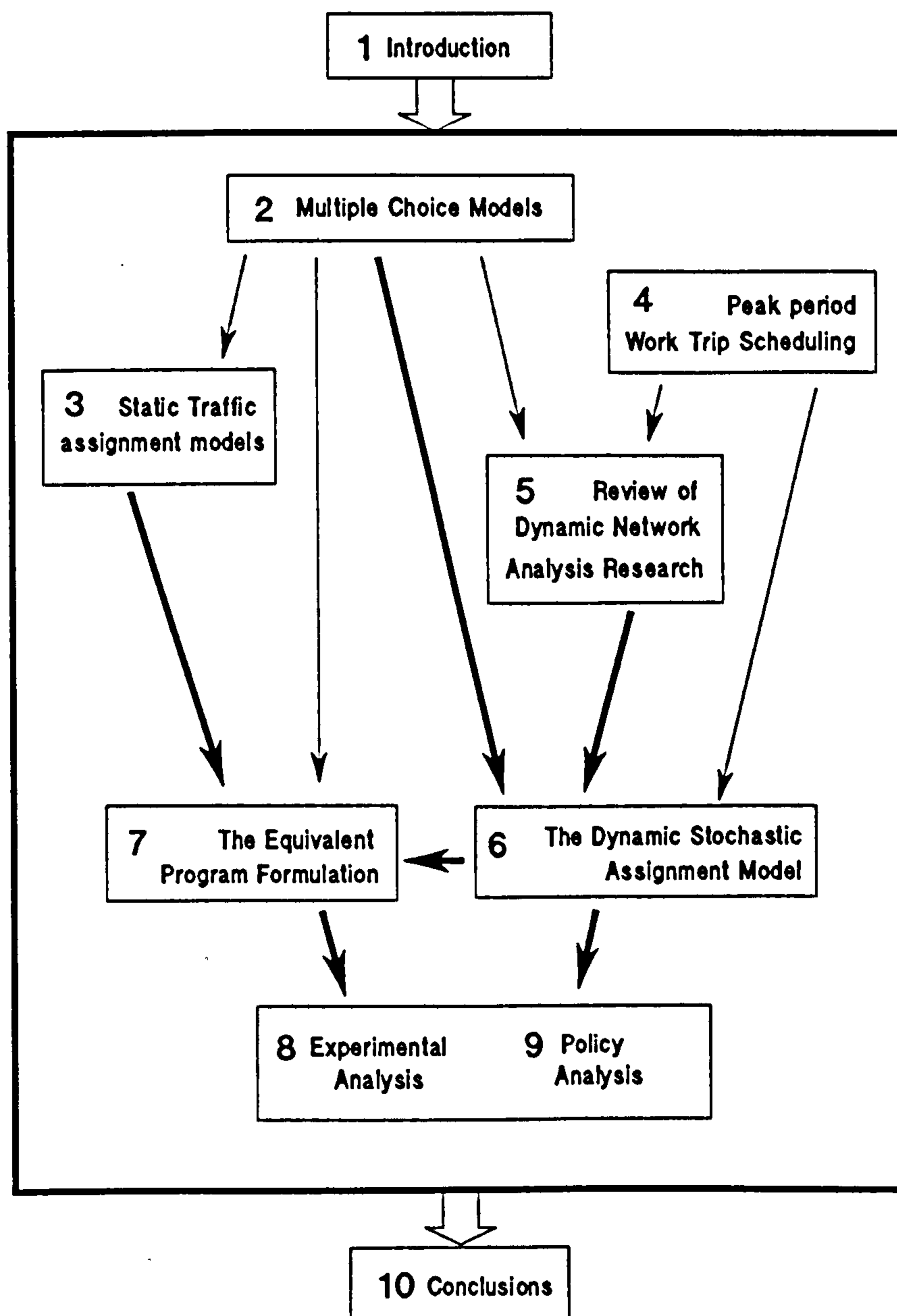


Figure 1.1 : chapter development

In chapter 6 the dynamic stochastic user equilibrium assignment model is developed. The model uses the theory presented in preceding chapters and further extends it in order to build a model that predicts travellers' route and departure time choices and estimates the resulting time varying traffic flow and travel time patterns for any network form. The main components of the model are : a demand model which is based on the utility maximisation principles, two alternative formulations of a travel time model (the first based on traffic flow and the second based on queueing theory), and a demand adjustment mechanism which describes the evolution of travel patterns from day to day and is used to derive the equilibrium solution.

Chapter 7 presents a framework for formulating and solving the dynamic assignment problem as a mathematical optimisation program. The major advantage of this approach is that it provides the framework of a methodology for solving the DSUE problem which does not require path enumeration. However, to achieve that, this framework requires a procedure which has not yet developed. Thus the algorithm used to derive the equilibrium flow patterns uses procedures developed in chapter 6 and which however require path enumeration.

Chapter 8 deals with the analysis of the results derived from the DSUE models developed in chapters 6 and 7. Numerical simulation experiments are conducted to analyse the impact that i) different work start time flexibilities, ii) different levels of demand, iii) different variability on perceptions of travel time and schedule delay and, iv) increases in a link capacity, have on the peak period traffic patterns.

Chapter 9 uses the developed DSUE assignment model as a policy analysis tool. Two transportation management policies are discussed. The first deals with the impacts of different road pricing policies on the formation of peak period traffic flow patterns and particular emphasis is given on the effects of such policies to departure time choice decisions. The second application deals with the evaluation of the effects of a route guidance system. The possible benefits to users and non-users of the system are evaluated under different percentages of the drivers who receive and use the information.

Finally chapter 10 summarises the work presented in this study. The major findings are discussed and their implications on transportation planning analysis. The chapter concludes with a discussion on the limitations of this research and the opportunities for further research.

## **2 multiple choice models**

### **Objective**

The purpose of this chapter is to present some of the currently available methods for modelling multiple choice problems which will be used to model route and departure time selection in chapters 6 and 7.

### **2.1 Introduction**

The principal requirement for a model of forecasting travel demand is that the model be *behavioural*, defined by Domencich and McFadden (1975) as one *which represents the decisions that consumers make when confronted with alternative choices*. Travellers are faced with a number of decisions such as whether to make a trip, where and when to go, which route to take, and which mode to select. Their decisions are based on their personal circumstances, (needs, income, occupation, car ownership, etc) and on the nature of the travel choices offered to them (travel time, costs, service levels of the alternatives, etc.). The model must therefore represent the relationships between the travellers' socioeconomic characteristics and the transport system attributes on the one hand, and the rate of demand for travel on the other.

A major subject of interest in the transportation planning process is the behaviour of a large number of individuals. However, aggregate behaviour is a result of individual decisions, and therefore the criteria and mechanisms leading an individual to a certain choice form the core of a demand forecasting model.

Choice was defined by Ben-Akiva and Lerman (1985) as

"an outcome of a sequential decision-making process that includes the following steps :

1. definition of the choice problem
2. generation of alternatives



3. evaluation of attributes of the alternatives
4. choice
5. implementation"

Thus a theory of choice is a collection of procedures that defines the following elements:

i. Decision maker

The decision maker may be an individual or a group of persons. Individuals may face different choice situations and have widely different requirements and tastes.

ii. Alternatives

Any choice is made from a set of alternatives. The environment of the decision maker determines the *universal set* of alternatives. Any single decision maker considers a subset of this universal set, the *choice set*. The latter set includes the alternatives that are both feasible to the decision maker and known during the decision process.

iii. Attributes of alternatives

The attractiveness of an alternative is evaluated in terms of a vector of attribute values which are measured on a scale of attractiveness.

iv. Decision rule

The decision rule describes the internal mechanisms used by the decision maker to process the available information and arrive at a unique choice. These rules can be classified into the following categories :

- *Dominance*. An alternative is dominant with respect to another if it is better for at least one attribute and no worse for all other attributes. This rule does not lead to a unique choice but can be used to eliminate inferior alternatives from a choice set.
- *Satisfaction*. For every attribute a level of satisfaction is assumed as a criterion for choice. Thus an alternative can be eliminated if it does not meet the criterion of at least one attribute. This rule alone may not necessarily lead to a choice but in combination with other rules such as dominance, it can be more decisive.
- *Utility*. The attractiveness of an alternative expressed by a vector of attributes values is reduced to a scalar. This defines a single objective function expressing the attraction of an alternative in terms of its attributes. In the following this index of attractiveness will be referred to as utility, a measure that the decision maker attempts to maximize through his or her choice

The modelling methodologies described in this chapter are associated with the concept of *utility maximisation* since this concept provides the basis of the *random utility*

theory, which is the most widely used choice theory in travel demand analysis, and will be employed in the modelling of dynamic assignment. Other choice theories such as the economic consumer theory and the constant utility theory are discussed by Ben-Akiva and Lerman (1985).

Therefore in the following section the concept of the concept of utility maximisation is introduced. Then the two most widely used discrete choice models, based on the framework of utility maximisation, namely the multinomial logit and probit models are presented in section 2.3 and 2.4 respectively. A discussion on the continuous and the nested logit model, used in modelling dynamic assignment in chapters 6 and 7, follows in sections 2.5 and 2.6. A dynamic extension of the multinomial logit model, applied in chapter 6 in order to estimate the temporal and stationary distributions of travel demand, is described in section 2.7. Section 2.8 introduces the concept of satisfaction and the satisfaction function, employed in the formulation of dynamic assignment as an optimization problem in chapter 7. Finally section 2.9 summarizes the chapter.

## 2.2 The Utility Maximisation Concept

The main hypothesis underlying the utility maximisation concept is that the attractiveness of a particular alternative  $i$  can be described with an *attractiveness* or *utility* measure associated with that alternative, denoted by  $U_i$ . The utility is a function of the attributes of the alternatives as well as the characteristics of the decision maker. Each individual then, is assumed to choose the alternative, which for him, has the highest value of  $U_i$ . However, utility cannot be measured directly since it is further associated with attributes and characteristics which may not be easily measured or even cannot be observed and therefore it must be treated as *random*. Consequently, the utilities are modelled as random variables, implying that the choice models only provide the *probability* that a certain alternative is selected.

Let  $U = (U_1, \dots, U_i, \dots, U_J)$  denote the vector of utilities associated with a given *choice set* of alternatives,  $C$ , consisting of  $J$  alternatives numbered  $1, 2, \dots, J$ . The utility is expressed as :

- a function of the observed attributes of the alternatives and the observed characteristics of the decision maker. Therefore it is expressed as  $U_i = U_i(\mathbf{a})$ , where  $\mathbf{a}$  denote the vector of variables which incorporate these attributes and characteristics, and also as :
- a random variable consisting of a deterministic component,  $V_i(\mathbf{a})$ , and a random

error  $e_i(\mathbf{a})$ , to include the random effects of the unobserved attributes of the alternative and characteristics of the traveller.

Therefore the utility function can be formulated as :

$$U_i(\mathbf{a}) = V_i(\mathbf{a}) + e_i(\mathbf{a}) \quad \forall i \in C \quad (2.1)$$

The random error satisfies the condition  $E[e_i(\mathbf{a})] = 0$ , implying that  $E[U_i(\mathbf{a})] = V_i(\mathbf{a})$

$U_i(\mathbf{a})$  is usually referred to as the *random* or *perceived utility*, since it reflects how decision makers perceive the attractiveness of a certain option as a result of perception errors due to lack of information on the attributes of the alternatives, irrational or probabilistic decision making and other factors (discussed by Williams (1977)).  $V_i(\mathbf{a})$  is usually referred to as the *systematic* or *measured utility*, since it captures the observed and measurable attributes of the alternatives and characteristics of the individuals.

The factors influencing the random error  $e_i$  are mainly unobservable and therefore the value of  $e_i$  can only be represented in terms of a probability distribution. The form of the joint probability distribution of  $e_i$  can be derived from observation of actual choices subject to the condition that the measured utility can be estimated. The most common functions used to represent the joint probability distribution of the error  $e_i$  are the logit and normal distributions. The most important characteristic of these joint distributions is the variance/covariance matrix, since the means of the random elements are defined as zero. This matrix is expressed as :

$$\Sigma = \begin{vmatrix} \sigma_1^2 & \sigma_{12} & \dots & \sigma_{1J} \\ \sigma_{21} & \sigma_2^2 & \dots & \sigma_{2J} \\ \cdot & \cdot & \dots & \cdot \\ \cdot & \cdot & \dots & \cdot \\ \cdot & \cdot & \dots & \cdot \\ \sigma_{J1} & \sigma_{J2} & \dots & \sigma_J^2 \end{vmatrix} \quad (2.2)$$

The diagonal elements  $\sigma_1^2 \dots \sigma_i^2 \dots \sigma_J^2$  represent the variances of the random elements  $e_i$ , and the off diagonal elements  $\sigma_{12} \dots \sigma_{ij} \dots$ , etc represent the covariances.

The probability that alternative  $i$  is chosen, denoted by  $P(i)$ , is the probability that the utility associated with this alternative is higher or equal to the utility of any other alternative of the choice set,  $C$ . It is assumed that  $C$  can be specified by using some reasonable deterministic rules† . Thus :

† Richardson (1982) and Swait and Ben Akiva (1987) review alternative approaches to the choice set generation problem.

$$\begin{aligned}
 P(i) &= \Pr(U_i \geq U_j, \forall j \in C, j \neq i) && \forall i \in C \\
 &= \Pr(V_i - V_j + e_i \geq e_j, \forall j \in C, j \neq i) && \forall i \in C
 \end{aligned}
 \tag{2.3}$$

Any multinomial choice model can be derived using equation (2.3) given the joint distribution of the error terms. Let  $f(e_1, e_2, \dots, e_j, \dots, e_k)$  denote the joint density function of the random errors; then  $P(i)$  can be defined as :

$$\begin{aligned}
 P(i) &= \int_{e_i=-\infty}^{+\infty} \int_{e_1=-\infty}^{V_i-V_1+e_i} \int_{e_2=-\infty}^{V_i-V_2+e_i} \dots \int_{e_{i-1}=-\infty}^{V_i-V_{i-1}+e_i} \int_{e_{i+1}=-\infty}^{V_{i+1}-V_1+e_i} \dots \dots \dots \\
 &\int_{e_j=-\infty}^{V_i-V_j+e_i} f(e_1, e_2, \dots, e_i, \dots, e_j) de_j de_{j-1} \dots de_{i+1} de_{i-1} \dots de_2 de_1 de_i
 \end{aligned}
 \tag{2.4}$$

Another more convenient way to express  $P(i)$  is to reduce the multinomial problem to a binary one. This is achieved since the condition

$$U_i \geq U_j \quad \forall j \in C, j \neq i$$

is equivalent to

$$U_i \geq \max_{\text{all } j \neq i} U_j \tag{2.5}$$

and therefore the probability of selecting  $i$  is expressed as :

$$P(i) = \Pr[ V_i + e_i \geq \max_{\text{all } j \neq i} (V_j + e_j) ] \tag{2.6}$$

The calculation of choice probabilities by utility maximisation is very complex when there are more than two alternatives and a solution in closed form can be derived only when certain assumptions are made about the joint distribution for the set of  $e_j$ . However as Langdon (1984a) argues, the calculation of  $P(i)$  is not too difficult if the distributions of  $e_j$  for each alternative are not correlated, implying that all the covariance terms are zero and thus the variance/covariance matrix (2.2) is diagonal. In this case equation (2.4) is equivalent to :

$$\Pr(V_i - V_j + e_i > e_j) = \int_{e_i=-\infty}^{+\infty} f_i(e_i) \left\{ \prod_{\text{all } i \neq j} \int_{e_j=-\infty}^{V_i+e_i-V_j} f_j(e_j) de_j \right\} de_i \quad \forall i \in C \tag{2.7}$$

where  $f_i(e_i)$  represents the probability distribution function of  $e_i$ .

In the following sections some choice models based on the utility maximisation principles will be presented. An extended analysis of these models, as well as of other utility maximisation based choice models which will not be used in the modelling of

dynamic assignment (and therefore are not discussed in this thesis), is presented by Ben-Akiva and Lerman (1985).

### 2.3 The Multinomial Probit

The multinomial probit (MNP) is a random utility model in which the error terms have a joint multivariate normal (MVN) distribution with zero mean and an arbitrary variance/covariance matrix,  $\Sigma$  (formulated in eq. (2.2)). The distribution of the error term vector can then be expressed as :

$$e \sim \text{MVN}(0, \Sigma) \quad (2.8)$$

where the density function of the MVN is given by :

$$f(e) = (2\pi)^{-J/2} |\Sigma|^{-1/2} \exp \left[ -\frac{1}{2} \cdot e \cdot \Sigma^{-1} \cdot e^T \right] \quad (2.9)$$

Given the variance/covariance matrix of the error terms and the vector of the measured utilities of the alternatives, the distribution of the perceived utility vector  $U$  can be modelled as multivariate normal :

$$U \sim \text{MVN}(V, \Sigma) \quad (2.10)$$

with density function :

$$f(U) = (2\pi)^{-J/2} |\Sigma|^{-1/2} \exp \left[ -\frac{1}{2} \cdot (U-V) \cdot \Sigma^{-1} \cdot (U-V)^T \right] \quad (2.11)$$

The MNP choice probabilities can then be estimated by substituting the joint density function of the error terms, (eq. (2.9)), in equation (2.4). However the form of the latter equations is such, that prohibits a closed form expression of the MNP choice probabilities. Thus, four choice probability approximation methods were developed to calculate these choice probabilities; these are : the numerical integration method, the tabulation method (interpolation between tabulated values of the cumulative MVN function), the numerical approximation, and the Monte Carlo simulation. However the first two (reviewed by Sheffi *et al.* (1982)) are impractical for problems involving more than three or four alternatives. So, in the remainder of this section only the numerical approximation and the simulation method will be discussed.

The numerical approximation technique for the calculation of the MNP choice probabilities was first suggested by Daganzo and Sheffi (1977) in the context of stochastic traffic assignment. This approximation is based on Clark's (1961) method of approximating the maximum of two normally distributed variables by a normal variate

and applying this approximation recursively. More specifically, Clark first considered 3 random variables  $U_i$ ,  $i=1,2,3$ , having an unrestricted joint normal distribution and obtained formulae for the exact values of the first four moments ( $n_1 \dots n_4$ ) of the maximum of  $U_1$  and  $U_2$ ,  $\max(U_1, U_2)$ , and the coefficient of linear correlation of  $\max(U_1, U_2)$  and  $U_3$ . He then approximated the distribution of the  $\max(U_1, U_2)$  by a normal distribution as :

$$\max(U_1, U_2) \sim N( n_1 , n_2 - n_1^2 ) \quad (2.12)$$

This approximation was then applied to derive the approximate distribution of 3 or more variables using the following recursive formula :

$$\max(U_1, U_2, \dots, U_{J-1}, U_J) = \max [ \max (U_1, U_2, \dots, U_{J-1}), U_J ] \quad (2.13)$$

Thus after  $J-1$  iterations, the maximum of the  $J$  variables can be approximated as :

$$\max(U_1, U_2, \dots, U_{J-1}, U_J) \sim N( V_{\max} , \sigma_{\max}^2 ) \quad (2.14)$$

If  $U_i$  is the last variable to be considered, then let denote

$$V_{\max(-i)} = E[ \max (U_1, \dots, U_{i-1}, U_{i+1}, \dots, U_J) ] \quad (2.15)$$

$$\sigma_{\max(-i)}^2 = \text{VAR}[ \max (U_1, \dots, U_{i-1}, U_{i+1}, \dots, U_J) ] \quad (2.16)$$

and

$$r_{-i,i} = \text{CORR}[ U_i , \max (U_1, \dots, U_{i-1}, U_{i+1}, \dots, U_J) ] \quad (2.17)$$

The probability that the  $i$ th variate is the largest, i.e. the probability that the  $i$ th alternative is chosen, is then given by :

$$\begin{aligned} P(i) &= \Pr \{ U_i \geq \max_{\text{all } j \neq i} U_j \} \\ &= \Pr \{ [\max(U_1, \dots, U_{i-1}, U_{i+1}, \dots, U_J)] - U_i \leq 0 \} \\ &= \Phi \left( \frac{V_i - V_{\max(-i)}}{\sigma_i^2 + \sigma_{\max(-i)}^2 - 2 \sigma_i \sigma_{\max(-i)} r_{-i,i}} \right) \end{aligned} \quad (2.18)$$

Due to the error introduced by equation (2.12), equation (2.18) is only an approximation. The accuracy of this approximation was investigated by Daganzo *et al.* (1977) who concluded that it results in unsatisfactory choice probabilities in the case where the alternatives have similar means (measured utilities) and very different variances. Horowitz *et al.* (1982) analysed some of the problems associated with Clark's

method and argues that the accuracy of the results depends greatly on the model of the utility function used. Langdon (1984b) has developed another approximation method which was proved to avoid the undesirable features of Clark's approximation at the expense of additional computational requirements, but it can be applied to cases with not more than 15 alternatives, and therefore is impractical for modelling dynamic assignment.

Another approach to compute the probit choice probabilities is the Monte Carlo simulation method suggested by Lerman and Manski (1978). Given the values of the measured utility  $V = (V_1, \dots, V_J)$  and the joint density function of the error terms  $e \sim MVN(0, \Sigma)$ , the simulation technique works iteratively as follows: At every iteration  $n$ , a vector  $e^n = (e_1^n, \dots, e_J^n)$  consisting of  $J$  random terms is drawn from the joint density function  $f(e)$ . Then the perceived utility of each alternative is defined as the sum of the perceived utility and the random term, that is  $U_i^n = V_i + e_i^n, \forall i$ . The alternative associated with the maximum perceived utility is then recorded. This procedure is repeated  $N$  times. Thus, if  $N_i$  denotes the number of times that each alternative  $i$  was recorded as being the one associated with the maximum perceived utility, the probability of selecting the  $i$ th alternative,  $P(i)$  is given by:

$$P(i) = \lim_{N \rightarrow \infty} \frac{N_i}{N} \quad (2.19)$$

$$\text{Thus, for sufficient large values of } N, P(i) \text{ is expressed as: } P(i) = \frac{N_i}{N} \quad (2.20)$$

For reasons of efficiency, Lerman and Manski (1978) suggested that the number of simulations,  $N$ , should not be fixed, but meet the following criterion:

$$N > \max_i \frac{1 - P(i)}{P(i) \cdot t_f} \quad (2.21)$$

where  $t_f$  is a tolerance level.

However as Daganzo *et al.* (1977) argues, this procedure introduces a bias in the estimate of  $P(i)$ , since this estimate is dependent on the numbers of simulations performed; this bias may be small for large probabilities but may be considerable for small probabilities. Furthermore Sheffi *et al.* (1982) comment that since the relative error associated with the simulation is inversely proportional to the square root of the number of successes, the computational cost associated with the simulation approach is considerable.

Despite the fact that the MNP provides, perhaps, the most satisfactory form of discrete choice model based on the framework of utility maximisation (Sheffi et al.(1982)), only a few applications of MNP have been appeared in the travel demand literature; the reason for its limited application is mainly the existence of the approximation errors and the high computational requirements associated with the already developed methods of evaluating MNP probability choices, which make it impractical for large scale problems. Furthermore as Ben-Akiva and Lerman (1985) stated "*there is still no evidence to suggest in which situations the greater-generality of multinomial probit is worth the additional computational problems resulting from its use*".

A complete analysis of the theory of MNP is presented by Daganzo (1979)

#### 2.4 The Multinomial Logit

The multinomial probit model presented in the previous section can manipulate any correlation between the alternatives, but is inefficient and impractical to use in large applications. Furthermore, as mentioned in section 2.2 the calculation of choice probabilities by utility maximisation is very complex when there are more than two alternatives and a solution in closed form can be derived only when certain assumptions are made about the distributions of the error terms.

For this reason certain assumptions were introduced to develop another model of multiple choice, termed the multinomial logit (MNL), which is widely used in transportation planning.

In the formulation of the MNL it is assumed that for all the alternatives  $i \in C$ , the random elements are :

- i. independently distributed,
- ii. identically distributed and
- iii. Gumbel-distributed with a location parameter  $n$ , and a scale parameter  $\mu > 0$ .

The assumption that the error terms are Gumbel distributed is used for reasons of analytical convenience as an approximation to normal distribution. The basic properties of the Gumbel distribution, used to derive the multinomial logit model, are presented in appendix A1. However, the assumption that the random terms are independent and identically distributed (IID) is rather restrictive. It implies that all the disturbances have the same scale parameter  $\mu$  and therefore that the variances of the random components of the utilities of all the alternatives are equal.



Using the properties of the Gumbel distribution, presented in appendix A1, the multinomial logit can be easily derived. The proof described below is the one suggested by Ben-Akiva and Lerman (1985).

For convenience it is assumed that  $n=0$  for all the random elements. Thus using equation (2.6)

$$P(i) = \Pr[ V_i + e_i \geq \max_{\text{all } j \neq i} (V_j + e_j) ] \quad (2.22)$$

Define

$$U^* = \max_{\text{all } j \neq i} (V_j + e_j) \quad (2.23)$$

Property 7 (appendix A1) implies that  $U^*$  is Gumbel distributed with parameters

$$\left( \frac{1}{\mu} \ln \sum_{\text{all } j \neq i} e^{\mu V_j}, \mu \right) \quad (2.24)$$

Using property 4, we can write  $U^* = V^* + e^*$ , where

$$V^* = \frac{1}{\mu} \ln \sum_{\text{all } j \neq i} e^{\mu V_j} \quad (2.25)$$

and  $e^*$  is Gumbel distributed with parameters  $(0, \mu)$ .

Since

$$\begin{aligned} P(i) &= \Pr( V_i + e_i \geq V^* + e^* ) \\ &= \Pr[ (V^* + e^*) - (V_i + e_i) \leq 0 ] \end{aligned} \quad (2.26)$$

using property 5

$$\begin{aligned} P(i) &= \frac{1}{1 + e^{\mu(V^* - V_i)}} = \frac{e^{\mu V_i}}{e^{\mu V_i} + e^{\mu V^*}} = \\ &= \frac{e^{\mu V_i}}{e^{\mu V_i} + \exp(\ln \sum_{\text{all } j \neq i} e^{\mu V_j})} = \frac{e^{\mu V_i}}{\sum_{\text{all } j \in C} e^{\mu V_j}} \end{aligned} \quad (2.27)$$

A different proof is presented by Domencich and McFadden (1975).

The scale parameter  $\mu$  defines the variance of the random elements (as they are assumed Gumbel distributed - property 3) and therefore reflect the variability of preferences among the decision makers. The presence of this parameter in each of the

terms of equation (2.27), implies two limiting cases of the MNL, both of which involve extreme values of this parameter.

The first case is for  $\mu \rightarrow 0$ , implying that the variance of the errors approach infinity. The choice model then provides no information and all the alternatives are equally likely.

$$\lim_{\mu \rightarrow 0} P(i) = \frac{1}{k} \quad \forall i \in C \quad (2.28)$$

For the second case  $\mu \rightarrow \infty$ , implying that the variance of the random elements approaches zero and thus a deterministic model is obtained since all the information about individuals' preferences is included in the systematic utilities.

$$\lim_{\mu \rightarrow \infty} P(i) = \lim_{\mu \rightarrow \infty} \frac{1}{1 + e^{\mu(V^* - V_i)}} = \begin{cases} 1 & \text{if } V_i \geq V^* \Leftrightarrow V_i \geq \max_{\text{all } j \neq i} V_j \\ 0 & \text{if } V_i \leq V^* \Leftrightarrow V_i \leq \max_{\text{all } j \neq i} V_j \end{cases} \quad (2.29)$$

$$= \begin{cases} 1 & \text{if } V_i \geq V^* \Leftrightarrow V_i \geq \max_{\text{all } j \neq i} V_j \\ 0 & \text{if } V_i \leq V^* \Leftrightarrow V_i \leq \max_{\text{all } j \neq i} V_j \end{cases} \quad (2.30)$$

In the occurrence of

$$V_i = \max_{\text{all } j \neq i} V_j$$

and there are  $k'$  equal alternatives which are also equal to  $\max_{\text{all } j \in C} V_j$ , then

$$P(i) = \frac{1}{k'} \quad (2.31)$$

A widely discussed aspect of the MNL is the independence of irrelevant alternatives (IIA) property implying that the ratio of the probabilities of choosing any two alternatives :

$$\frac{P(i)}{P(k)} = \frac{e^{\mu V_i} / \sum_{j \in C} e^{\mu V_j}}{e^{\mu V_k} / \sum_{j \in C} e^{\mu V_j}} = \frac{e^{\mu V_i}}{e^{\mu V_k}} \quad (2.32)$$

is independent of the attributes of any other alternatives. This property has some important consequences, since it can result in erroneous predictions. The reason for the problematic behaviour of the MNL is the assumption that the error terms are mutually independent, implying that the sources of the errors contributing to the random terms

must be such that the total random terms are independent. However this assumption is not justifiable in the case where at least two alternatives have common unobservable characteristics.

A way to show the implications of the IIA property in the case where the MNL is used to calculate choice probabilities when the alternatives are correlated, is to go through the steps of the MNL derivation. Thus let  $V_1, V_2, V_3$  denote the systematic utilities associated with alternatives 1,2,3, where alternatives 2 and 3 have identical observed attributes and therefore  $V_2=V_3=V$ . Then from equations (2.24) and (2.25)

$$U^* = \max(U_2, U_3) = V^* + e^* = \frac{1}{\mu} \ln(e^{\mu V_2} + e^{\mu V_3}) + e^* = \ln 2 / \mu + V + e^* \quad (2.33)$$

However if the alternatives are perfectly correlated, then, since the systematic utilities of 2 and 3 are equal, the maximum of  $U_2$  and  $U_3$  is  $V + e$ . Thus the IIA assumption results in an overestimation of  $V^*$  by  $\ln 2 / \mu$  and therefore in an inaccurate estimation of the choice probabilities. These probabilities as calculated assuming that :

- i) the IIA property holds for all the alternatives and
- ii) a perfect correlation exists between alternatives 2 and 3

are the following :

i) IIA

$$V^* = \ln 2 + V$$

$$P(1) = \frac{1}{1 + 2 e^{\mu(V-V_1)}}$$

$$P(2) = P(3) = \frac{e^{\mu(V-V_1)}}{0.5 + e^{\mu(V-V_1)}}$$

ii) Correlation

$$V^* = V$$

$$P(1) = \frac{1}{1 + e^{\mu(V-V_1)}}$$

$$P(2) = P(3) = \frac{e^{\mu(V-V_1)}}{1 + e^{\mu(V-V_1)}}$$

The above equations show that application of the MNL in the case that alternatives 2 and 3 are perfectly correlated results in an underestimation of the choice probability for alternative 1 and overestimation for alternatives 2 and 3.

The IIA property is discussed by several researchers, including McFadden *et al.* (1977) and Ben-Akiva and Lerman (1985) who also analyse and other properties of the MNL.

Despite its limitations the MNL is the most widely used model for estimating choice probabilities in transportation-demand analysis since, as was stated by McFadden *et al.* (1977), "has significant advantages over the available alternatives in terms of flexibility and computational efficiency and permits a simple behavioural interpretation of the parameters of the scale (utility) function".

In the following sections some developments of the MNL, which will be used in subsequent chapters for modelling dynamic assignment, will be presented.

### 2.5 The Continuous Logit

A common assumption in all the models already presented in this chapter was that the alternatives consisting the choice set are discontinuous. However individuals are often faced with choice decisions where the alternatives are *continuous variables* formulating a *continuous choice set*.

The continuous logit model has a form similar to the multinomial logit, where the denominator is modified so that to be an integral over the continuous choice set, rather than a sum as in the MNL, and is defined as:

$$P(i) = \frac{e^{\mu V(i)}}{\int_H e^{\mu V(h)} dh} \quad (2.34)$$

where the measured utility from alternative  $h$ , is expressed as a direct function of  $h$ , and the integration is performed over the continuous choice set  $H$ .

The model was first developed by McFadden (1976), who assumed that the IIA property holds with respect to subsets of the continuous alternatives. Ben-Akiva and Watanada (1981) applied the continuous logit model in the context of spatial choice; their model was expressed in terms of two-dimensional coordinates to represent the location of the spatial alternatives. Ben-Akiva *et al.*(1985) derived the continuous logit model directly from the random utility assumption, for a single and a conditional choice case. They showed that the continuous logit model is a generalization of the discrete logit model and applied it to derive spatial distributions of residential locations, workplaces and trips.

The continuous logit was initially used to model the choice of departure time in the morning work trip by Ben-Akiva *et al.*(1984). They formulated the observed utility as a function of the departure time from the origin, denoted by  $V(t)$ , and expressed the probability of time  $t$  being selected, by the following continuous logit model :

$$P(t) = \frac{e^{\mu V(t)}}{\int_T^{T+T_0} e^{\mu V(u)} du} \quad (2.35)$$

where all drivers are assumed to depart from origin between  $T$  and  $T+T_0$ .

## 2.6 The Nested Logit

In the forms of the logit model analysed in the preceding sections, the choice set considered, was relatively simple. However there are situations where *the members of the set of feasible alternatives are combinations of underlying choice dimensions* (Ben-Akiva and Lerman (1985)). For example, in modelling commuting trips each alternative for a driver, travelling from origin  $r$ , might be defined by both the departure time and the route used to reach his destination,  $s$ .

In this section a generalisation of the logit model, termed the *nested logit* (NL), is presented, which is used in cases where the choice set is multidimensional. The NL was designed to improve the MNL since the latter, due to the IIA property inherent in its formulation, is rather inappropriate to apply when the alternatives are correlated. In the NL, as in the MNL, every individual is assumed to evaluate each of the alternatives according to their utility functions. He is also assumed to decompose his trip into several stages, establish a *hierarchy* between these stages, and then follow a *sequential* decision making process, as shown in figure (2.1). Thus the alternatives which are correlated are grouped together in *clusters* or *nests*, and represented by an aggregate alternative with a *composite* utility.

The concept of the *hierarchical structure* of the alternatives was introduced by Ben-Akiva (1973) who studied the choices of mode and destination for shopping trips in Washington D.C. The NL was subsequently formulated by several researchers including Williams (1977), Daly and Zachary (1978) and Ben-Akiva and Lerman (1979).

In the remainder of this section the assumptions and derivation of the NL is presented based on the analysis by Ben-Akiva and Lerman (1985) and using the route and departure time example mentioned earlier. Thus in this two-dimensional choice problem, (which will be extended later in multiple dimensions) two choice sets are considered :

$R = \{r_1, r_2, \dots, r_n\} = \{ \text{all possible routes connecting the O/D pair } r/s \}$

$T = \{t_1, t_2, \dots, t_k\} = \{ \text{all possible departure times from } r \}$

Thus :

$R \times T = \{ (r_1, t_1), (r_1, t_2), \dots, (r_1, t_k), (r_2, t_1), \dots, (r_2, t_k), \dots, (r_n, t_1), \dots, (r_n, t_k) \}$

will be all potential route and departure time combinations.

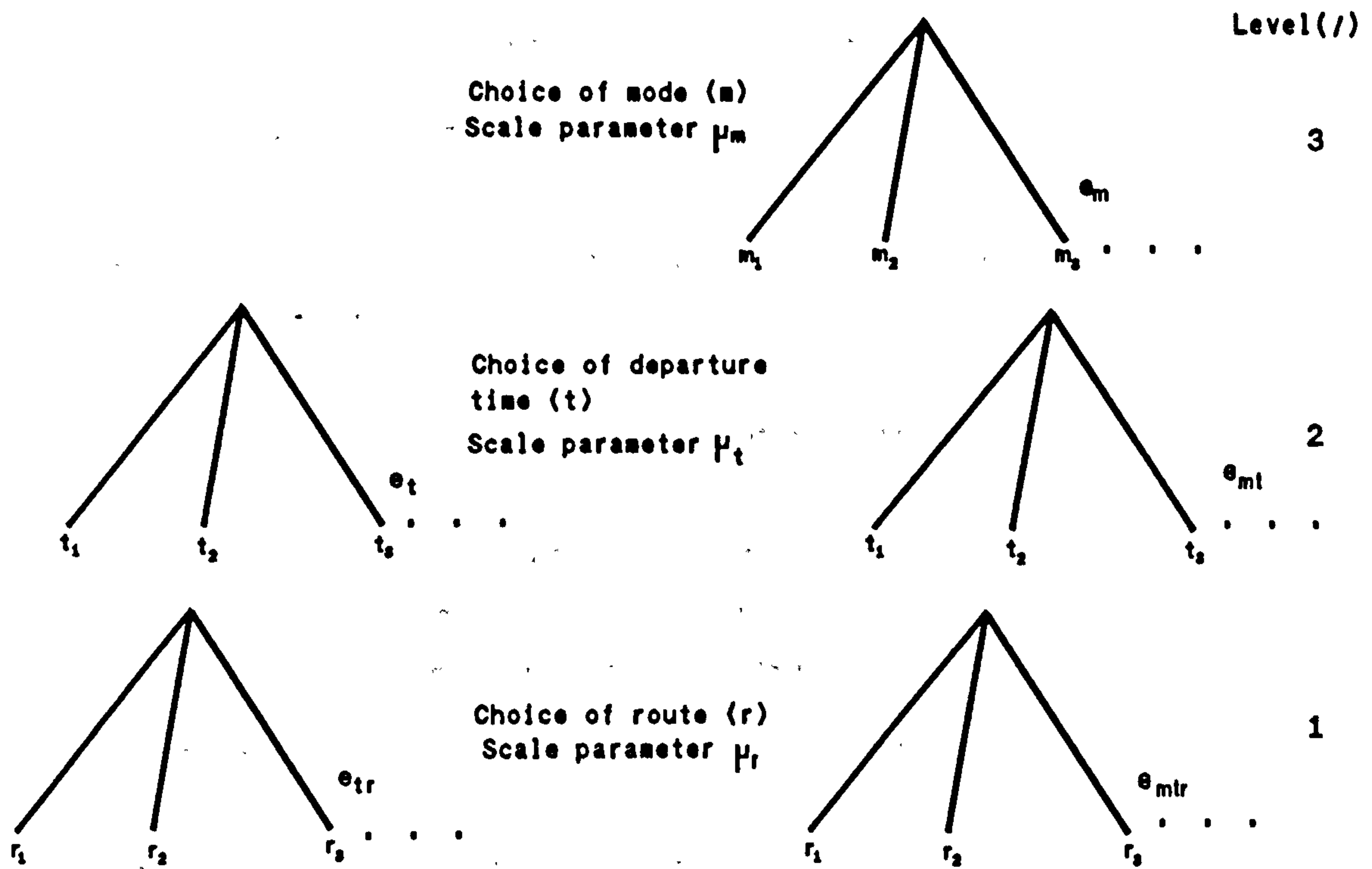


Figure 2.1 :  
Hierarchical Structure of the Nested Logit for two and three dimensions

The hierarchical structure of the alternatives and the sequential decision making process assumed, implies that an individual will first have to decide which of the alternatives at the upper level ( $l=2$  in fig.(2.1)) of the hierarchy to select, that is at what time  $t \in T$  to depart, and then subject to his choice, to select which route to follow. However, all the routes  $r \in R$  may not be feasible for a driver departing at time  $t$  (e.g. some of the routes may be rejected as being highly congested), and therefore for each departure time  $t$ , a new choice set is formulated, termed the *conditional* route choice set, which is defined as a subset of  $R$  including the routes that are feasible to an individual who departs at time  $t$  and is denoted by  $R_t$ . Thus the probability that an individual will select the combination of departure time  $t$  and route  $r$  can be expressed as :

$$P(t,r) = P(r|t)P(t) \tag{2.36}$$

where  $P(r|t)$  denotes the conditional probability of selecting route  $r$  subject to the choice of departure time  $t$ .

Before proceeding in the formulation of the above probabilities, the form of the utility functions should be specified. Thus, let  $U_{tr}$  denote the total utility perceived by a driver who selects route  $r$  and departs from origin at time  $t$ , expressed as:

$$U_{tr} = V_r + V_t + V_{tr} + e_r + e_t + e_{tr} \quad (2.37)$$

where

$V_t$  = the systematic component of utility common to all elements of  $C$  using departure time  $t$

$V_r$  = the systematic component of utility common to all elements of  $C$  using route  $r$

$V_{tr}$  = the remaining systematic component of utility specific to the combination  $(t,r)$

$e_t$  = the random element attributable to departure time  $t$

$e_r$  = the random element attributable to route  $r$

$e_{tr}$  = the remaining random element attributable to the combination  $(t,r)$

The significance of including the terms  $e_r$  and  $e_t$  in the utility function can be shown as follows.

Consider any two feasible alternative combinations of departure time and route, sharing a common departure time. The covariance of their utilities, assuming that each component of the random element is independent of all the other components, is expressed as :

$$\begin{aligned} \text{cov}(U_{tr}, U_{tr'}) &= \text{cov}(e_t + e_r + e_{tr}, e_t + e_{r'} + e_{tr'}) = \\ &= E[e_t^2] = \text{var}(e_t) \end{aligned} \quad (2.38)$$

Similarly it is shown that

$$\text{cov}(U_{tr}, U_{t'r}) = \text{var}(e_r) \quad (2.39)$$

Equations (2.38) and (2.39) imply that *when there are shared random components associated with different choice dimensions, the utilities of the elements of the corresponding multidimensional choice set cannot be independent.* However if the magnitudes of  $e_r$  and  $e_t$  are large compared to  $e_{tr}$  the resulting choice model may have a very complicated form. Thus, in order to cope with this problem, a main assumption of the NL is that  $e_r$  or  $e_t$  is small enough in magnitude so it can be reasonably ignored. Under this assumption, (e.g.  $e_r = 0$ ) the new form of the utility function is :

$$U_{tr} = V_r + V_t + V_{tr} + e_t + e_{tr} \quad (2.40)$$

Further assumptions in the formulation of the NL are :

1.  $e_t$  and  $e_{tr}$  are independent for all  $t \in T$  and  $r \in R$ .
2. The terms  $e_{tr}$  are independent and identically distributed with scale parameter  $\mu_r$ .
3.  $e_t$  is distributed so that  $\max_{r \in R_t} U_{rt}$  is Gumbel distributed with scale parameter  $\mu_t$ .

Thus following the random utility statements the probability of selecting departure time  $t$  is expressed as :

$$\begin{aligned}
 P(t) &= \Pr \left[ \max_{r \in R_t} U_{tr} \geq \max_{r \in R_{t'}} U_{t'r}, \forall t' \in T, t' \neq t \right] \\
 &= \Pr \left[ V_t + e_t + \max_{r \in R_t} (V_r + V_{tr} + e_{tr}) \geq \right. \\
 &\quad \left. \geq V_{t'} + e_{t'} + \max_{r \in R_{t'}} (V_r + V_{t'r} + e_{t'r}), \forall t' \in T, t' \neq t \right] \quad (2.41)
 \end{aligned}$$

Since  $e_{tr}$  is Gumbel distributed with parameter  $\mu_r$ , then

$$\max_{r \in R_t} (V_r + V_{tr} + e_{tr}) \quad (2.42)$$

is also Gumbel distributed with parameters

$$\left( \frac{1}{\mu_r} \ln \sum_{r \in R_t} e^{(V_r + V_{tr})\mu_r}, \mu_r \right) \quad (2.43)$$

Therefore eq.(2.41) can be expressed as :

$$P(t) = \Pr \left[ V_t + V_t^* + e_t + e_t^* > V_{t'} + V_{t'}^* + e_{t'} + e_{t'}^*, \forall t' \in T, t' \neq t \right] \quad (2.44)$$

where

$$V_t^* = \frac{1}{\mu_r} \ln \sum_{r \in R_t} e^{(V_r + V_{tr})\mu_r} \quad (2.45)$$

and expresses the expected maximum utility as defined by eq. (2.42)

$$e_t^* = \max_{r \in R_t} (V_r + V_{tr} + e_{tr}) - V_t^*, \quad (2.46)$$

is the new error term which is Gumbel distributed with parameter  $\mu_r$ .

The combined term  $e_t + e_t^*$  is, according to assumption 3, independent and identically Gumbel distributed with a scale parameter  $\mu_t$ , for all  $t \in T$ , and thus :

$$P(t) = \frac{e^{(V_t + V_t^*)\mu_t}}{\sum_{t' \in T} e^{(V_{t'} + V_{t'}^*)\mu_t}} \quad (2.47)$$



A necessary condition which must be held in order the NL to be valid is :

$$\frac{\mu_t}{\mu_r} \leq 1 \quad (2.48)$$

derived since, the variance of a Gumbel variate is inversely proportional to the square of its scale parameter. (Gumbel distribution property 3), and therefore

$$\begin{aligned} \frac{\mu_t}{\mu_r} &= \left[ \frac{\text{var}(e_{tr})}{\text{var}(e_t + e_t^*)} \right]^{1/2} \\ &= \left[ \frac{\text{var}(e_{tr})}{\text{var}(e_t) + \text{var}(e_{tr})} \right]^{1/2} \end{aligned} \quad (2.49)$$

since  $\text{var}(e_t^*) = \text{var}(e_{tr})$  as the variables  $e_t^*$  and  $e_{tr}$  have the same scale parameter, and it was assumed that  $\text{cov}(e_t^*, e_{tr}) = 0$  (assumption 1). Furthermore  $\mu_t/\mu_r = 1$  only if the variance of  $e_t$  is zero

The conditional choice probability for the nested logit model is expressed as:

$$\begin{aligned} P(r|t) &= \Pr [ U_{tr} \geq U_{tr'}, \forall r' \in R_t, r' \neq r | t \text{ chosen} ] \\ &= \Pr [ V_{tr} + V_r + e_{tr} \geq V_{tr'} + V_{r'} + e_{tr'}, \forall r' \in R_t, r' \neq r | t \text{ chosen} ] \end{aligned} \quad (2.50)$$

The components of the total utility due to  $V_t$  and  $e_t$  were omitted since they are constant for all the alternatives in  $R_t$ . Furthermore, since  $e_{tr}$  satisfies the assumptions of the MNL the conditional choice probability can be defined as :

$$P(r|t) = \frac{e^{(V_{tr} + V_r)\mu_t}}{\sum_{r' \in R_t} e^{(V_{tr'} + V_{r'})\mu_t}} \quad (2.51)$$

The probability of selecting the combination (t,r) can then be calculated from equations (2.36), (2.45), (2.47) and (2.51).

The two dimensional NL analysed above can be further extended to higher dimensions. Taking for a example the 3 dimensions, by adding the choice of mode (m) of travel for the commuting trip, the utility can be expressed as :

$$U_{mtr} = V_m + V_t + V_r + V_{mr} + V_{mt} + V_{tr} + V_{mtr} +$$

$$+ e_m + e_t + e_r + e_{mr} + e_{mt} + e_{tr} + e_{mtr} \quad (2.52)$$

For the nested model to be used the necessary assumptions as were stated by Ben-Akiva and Lerman (1985) are :

1. For all the levels of the hierarchy of choice, all components of the total random element involving level  $l$ , except the one involving also all the higher levels, have zero variance.

2. All the random terms are mutually independent.

3. For the levels  $l$ , the sum of the error terms at level  $l$  and those at the next lower level are identically Gumbel distributed.

Thus assumption 1 implies for  $l=1$  that  $\text{var}(e_r)=\text{var}(e_{tr})=\text{var}(e_{mr})=0$  and for  $l=2$  that  $\text{var}(e_t)=0$ . Assumption 3 implies that  $e_{mtr}$ ,  $(e_{mt}+e_{mtr})$  and  $(e_m+e_{mt}+e_{mtr})$  are Gumbel distributed. The decision tree for this example is illustrated in figure 2.1, where the lowest level ( $l=1$ ) is associated with the route choice and the highest ( $l=3$ ) with the mode choice. On each branch of the tree the relevant error term of the total utility is depicted. Under the preceding assumptions and denoting by  $\mu_r$  the scale parameter for  $e_{mtr}$ , by  $\mu_t$  the one for  $(e_{mtr} + e_{mt})$  and by  $\mu_m$  the one for  $(e_{mtr} + e_{mt} + e_m)$  the nested logit is expressed as :

$$P(mtr)=P(r|mt)P(t|m)P(m) \quad (2.53)$$

where

$$P(r|mt) = \frac{e^{(V_r+V_{tr}+V_{mr}+V_{mtr})\mu_r}}{\sum_{r' \in R_{mt}} e^{(V_{r'}+V_{tr}+V_{mr}+V_{mtr})\mu_r}} \quad (2.54)$$

$$P(t|m) = \frac{e^{(V_t+V_{mt}+V_{mt}^*)\mu_t}}{\sum_{t' \in T_m} e^{(V_{t'}+V_{mt}+V_{mt}^*)\mu_t}} \quad (2.55)$$

$$P(m) = \frac{e^{(V_m+V_m^*)\mu_m}}{\sum_{m' \in M} e^{(V_{m'}+V_m^*)\mu_m}} \quad (2.56)$$

and where

$$V_{mt}^* = \frac{1}{\mu_r} \ln \sum_{r \in R_{mt}} e^{(V_r+V_{tr}+V_{mr}+V_{mtr})\mu_r} \quad (2.57)$$

$$V_m^* = \frac{1}{\mu_t} \ln \sum_{t \in T_m} e^{(V_t+V_{mt}+V_{mt}^*)\mu_t} \quad (2.58)$$

As in the two dimensional case, the ratios  $\mu_t/\mu_r$  and  $\mu_m/\mu_t$  both must be positive and be less than or equal to 1 and therefore must satisfy the condition

$$\mu_m \leq \mu_t \leq \mu_r \quad (2.59)$$

In other words the variance of the random utilities is the smallest at the lower level of the tree and cannot decrease as we move from a low level to a higher level.

The NL was used by several researchers, including Sobel (1980), Hartley and Ortuzar(1980), Ortuzar (1983) and others, mainly in the context of modal split. It was first used by Ben-Akiva *et al.* (1986a,b) for modelling dynamic assignment in a network where drivers have to travel between one O/D pair connected with a number of parallel (not overlapping) routes but also have the option to switch to a different mode, with utility  $V_o$ . They expressed the probability that an individual will drive to work, depart at time  $t$  and select route  $i$ , as :

$$P(i,t) = \frac{e^{V^i(t)\mu_r}}{\sum_i e^{V^i(t)\mu_r}} \cdot \frac{e^{V^*(t)\mu_t}}{e^{V^*(*)\mu_t}} \cdot \frac{e^{V^*(*)\mu_m}}{e^{V^*(*)\mu_m} + e^{V_o\mu_m}} \quad (2.60)$$

where

$$V^*(t) = \frac{1}{\mu_r} \ln \sum_j e^{(V^j(t))\mu_r} \quad (2.61)$$

$$V^*(*) = \frac{1}{\mu_t} \ln \int_{T_o}^{T_o+T} e^{(V^*(u))\mu_t} du \quad (2.62)$$

The above choice probability is derived from the general form of the NL, eq.(2.53)...(2.56), assuming that the utility (eq. (2.52)) function has the form :

$$U_{mtr} = V_m + V_{tr} + e_m + e_{mt} + e_{mtr} \quad (2.63)$$

## 2.7 Dynamic Models of Choice

The term dynamic choice models is used to define the family of models characterised by the assumption that choice decisions are not static but evolve over time.

In the context of transportation, dynamic choice modelling was initially introduced by Hartgen (1974) who developed a dynamic model of mode choice using the adoption-

diffusion approach† to analyse the demand aspects of mode-switching behaviour when time dependent changes in usage occur.

In this section a *dynamic extension* of the MNL is presented. It is based on the work by de Palma and Lefevre (1983), who argues that " *the MNL is limited in its usefulness because it doesn't account for the effects of the time factor and the social interactions between individuals* ".

The new variable introduced in the dynamic models, is the time factor which is incorporated in such models in order to reflect the individual's attitude towards choice decisions which are not static but reviewed and evolve over time. Such a behaviour is a result of changes in the tastes or socio-economic characteristics of the individuals, and also of the fact that the attributes of a certain alternative usually depend on the choices made by other individuals having the same set of alternatives. Thus for example, the selection of route and departure time is crucially dependent on the choices made by other users of a transportation network, since travel time, a main component of the utility function, depends on the level of demand using the network.

The extension of the MNL that takes account of the above mentioned aspects of the problem is formulated as an interactive continuous-time Markov process.

Thus consider a population of  $n$  individuals, who have a choice set consisting of  $J$  alternatives  $j=1, \dots, J$ . Let  $Y_j(t)$ ,  $j=1, \dots, J$ ,  $t \geq 0$ , denote the number of individuals who choose alternative  $j$  at time  $t$ , and  $y_j$  a possible realization of  $Y_j(t)$ , where  $Y_1(t) + \dots + Y_J(t) = n$ . Let  $Y(t)$  be the column vector  $[Y_1(t), \dots, Y_J(t)]'$  and  $y$  a possible realization of  $Y(t)$ . The probability that an individual, during the time interval  $(t, t + \Delta t)$ , decides to review his actual choice  $i$  given the choice distribution of the population  $Y(t)=y$ , is defined as  $R_i(y,t)\Delta t + o(\Delta t)$ , and it is assumed that the probability of more than one review during  $(t, t + \Delta t)$  is of the order of  $o(\Delta t)$ . Once an individual decides to review his choice, he selects the alternative  $j$  which for him has the highest utility. Therefore the probability that he selects alternative  $j$  is given by the following MNL :

$$P_{j|i}(y,t) = \frac{e^{\mu V_{ji}(y,t)}}{\sum_{s=1}^J e^{\mu V_{si}(y,t)}} \quad j=1, \dots, J \quad (2.64)$$

where

$V_{ji}(y,t)$  represents the measured utility of a transition from alternative  $i$  to

---

† Choice can be treated as an adoption process in which the individuals who previously used another alternative begin to use a "new" alternative. As the number of persons *adopting* the new alternative increases, its use is said to *diffuse* through the population. Later as users become dissatisfied with the new alternative, or as other other alternatives become available, usage may fall, eventually reaching a steady-state level.

alternative  $j$  at time  $t$ , when the distribution of demand among the different alternatives is given by  $y$ .

The transition rates from  $i$  to  $j$  are then given as  $R_i(y,t)p_{ji}(y,t)$

Furthermore, let  $P_{y_0,y}(t)$  denote the probability that at time  $t$ , the global choice distribution of the  $n$  individuals is given by  $Y(t) = y$ , given that at time 0, it was given by  $Y(0) = y_0$ . Let also  $h_j = [\delta_{1j}, \dots, \delta_{Jj}]'$ ,  $j=1, \dots, J$ , be an orthonormal base of  $R^J$ . The marginal probability  $P_{y_0,y}(t)$  with respect to time is then given by the following Chapman-Kolmogorov equations :

$$\begin{aligned} \frac{dP_{y_0,y}(t)}{dt} = & \sum_{j|y_j \geq 1} \sum_{\substack{i|y_i \leq n-1 \\ i \neq j}} P_{y_0, y+h_i-h_j}(t) (y_i+1) R_i(y+h_i-h_j, t) p_{ji}(y+h_i-h_j, t) \\ & - P_{y_0,y}(t) \sum_{i|y_i \geq 1} y_i R_i(y, t) \sum_{j|j \neq i} p_{ji}(y, t) \end{aligned} \quad (2.65)$$

These are the master equations of a multivariate birth and death process. However an explicit solution of (2.65) cannot be derived.

A tractable approximation for such a Markov process, when the population size  $N$  is large, is presented by de Palma and Lefevre (1983) ; the basic element of the approximation is a deterministic process which is, in some sense, associated with the stochastic process.

Thus let  $Z^n(t)$  denote the column vector of the  $J$  choice densities  $Z_j^n(t) = Y_j(t)/n$ ,  $j=1, \dots, J$ , and  $Z^n(t) = Y(t)/n = [Z_1^n(t), \dots, Z_J^n(t)]'$ . Let  $z^n$  be a possible realisation of  $Z^n(t)$ , with  $Z^n(0) = z_0^n$ . A crucial hypothesis is that the transition rates depend on the global choice distribution precisely through the densities  $z^n$  (instead of the sizes  $z$  as assumed previously). Then using (2.65) the marginal probability that  $Z^n(t) = z^n$  given that  $Z^n(0) = z_0^n$ , denoted by  $P_{z_0^n, z^n}(t)$ , with respect to  $t$ , is expressed as :

$$\begin{aligned} \frac{dP_{z_0^n, z^n}(t)}{dt} = & \sum_{j|z_j^n \geq 1/n} \sum_{\substack{i|z_i^n \leq 1-1/n \\ i \neq j}} P_{z_0^n, z^n+h_i/n-h_j/n}(t) n (z_i^n+1/n) R_i(z^n+h_i/n-h_j/n, t) \\ & p_{ji}(z^n+h_i/n-h_j/n) - P_{z_0^n, z^n}(t) \sum_{i|z_i^n \geq 1/n} n z_i^n R_i(z^n) \sum_{j|j \neq i} p_{ji}(z^n) \end{aligned} \quad (2.66)$$

Then according to Kurtz (1978) the deterministic version of the Markov process (2.66) is the following system of differential equations :

$$\frac{dZ_j(t)}{dt} = \sum_{\substack{i=1 \\ \neq j}}^J Z_i(t) R_i[Z(t)] p_{ji}[Z(t)] - Z_j(t) R_j[Z(t)] \sum_{\substack{i=1 \\ \neq j}}^J p_{ij}[Z(t)] \quad (2.67)$$

This deterministic version was used by Ben-Akiva and de Palma (1986) to analyse a dynamic residential location choice model with transaction costs.

The system of differential equations (2.67) is very general. A simplifying assumption, that is the transition rates to be independent of the initial choice implies that the system (2.67) becomes :

$$\frac{dZ_j(t)}{dt} = R \{ p_j[Z(t)] - Z_j(t) \} \quad j = 1, \dots, J \quad (2.68)$$

This deterministic model has been used by Deneubourg *et al.* (1979) in studying the dynamics of choice of transportation mode. A simplified version of this model was first used in modelling dynamic assignment by Ben-Akiva *et al.* (1984), who expressed the rate of change of the number of individuals (travelling in a simple network consisting of a single O/D pair connected by one route) departing from the origin at time  $t$  during the time interval  $[w, w + \Delta w]$  as :

$$\frac{\partial Q(t, w)}{\partial w} = R \cdot \left[ n \frac{e^{mV(t, w)}}{\int_T^{T+T_0} e^{mV(u, w)} du} - Q(t, w) \right] \quad (2.69)$$

where

- $Q(t, w)$  is the departure rate from the origin at time  $t$  on day  $w$
- $V(t, w)$  is the measured utility for a driver departing at time  $t$  on day  $w$
- $T, T+T_0$  are the earliest and latest departure times from the origin
- $R$  is a constant transition rate out of the current state
- $n$  is the total demand

This model form was also used to predict the temporal and stationary distributions of travel demand for more complicated network and will be further analysed in chapters 5 and 6.

## 2.8 Satisfaction and the Expected Disutility of Travel

The focus of the previous sections of this chapter was on different multiple choice models and the computation of the choice probability. Another important quantity related to multiple choice models is the *satisfaction*,  $S$ , and the *satisfaction function*,

$S(\mathbf{a})$ . The concept of satisfaction in multiple choice models was suggested by Daganzo (1979), and also developed as an *accessibility* measure of the choice set  $C$  to a decision maker by Ben-Akiva and Lerman (1979). The satisfaction of any decision maker is the perceived utility from his choice. But since by definition of the utility maximisation concept, the perceived utility of the individual who selects a certain alternative is the highest, his satisfaction is given by the  $\max(U_j)$ ,  $j=1, \dots, J$ . Thus the expected satisfaction that a randomly selected decision maker derives from the choice set of alternatives, called satisfaction, is defined as the expected value of the maximum perceived utility of the alternatives within the choice set, and expressed as:

$$S = E[ \max_{j \in C} \{U_j\} ] \quad (2.70)$$

The function relating the satisfaction to the vector of the alternatives' attributes and the decision makers' characteristics,  $\mathbf{a}$ , is called the satisfaction function. Given that  $V = V(\mathbf{a})$  and given the distribution of the random errors  $e$ , the satisfaction is expressed as a function of  $V$ . Thus :

$$S(V) = E[ \max_{j \in C} \{V_j + e_j\} ] \quad (2.71)$$

The satisfaction function,  $S(V)$ , has three important properties :

- i. It is convex with respect to  $V$ .
- ii. The partial derivative of the satisfaction function with respect to the measured utility of an alternative equals its probability of choice :

$$\frac{\partial S(V)}{\partial V_i} = P(i) \quad (2.72)$$

- iii. The satisfaction is monotonic with respect to the size of the choice set :

$$S(V_1, V_2, \dots, V_J, V_{J+1}) \geq S(V_1, V_2, \dots, V_J) \quad (2.73)$$

Proof of the above properties are presented by Daganzo (1979). The second property, stating that the marginal expected maximum perceived utility with respect to a certain alternative is equal to the probability of choosing that alternative, was also proved by Ben-Akiva and Lerman (1979) and by Sheffi and Daganzo (1978) in the context of networks. This property holds only under the condition that the shape of the utilities density function does not depend on the measured utilities, while the third property holds only for non-interacting alternatives (i.e. if the introduction of the extra alternative does not decrease the measured utility of any other alternative).

For the logit models, the form of the satisfaction function is derived directly from its definition (eq. (2.71) ) and is given by a form similar to eq. (2.25) as :

$$S(V) = \frac{1}{\mu} \ln \sum_{\text{all } j \in C} e^{\mu V_j} \quad (2.74)$$

Having defined the satisfaction and satisfaction function, the concept of expected perceived *disutility* of travel is introduced in the remainder of this section, based on the derivation and properties of the expected perceived travel time analysed by Sheffi (1985). Disutility is the negative utility and reflects the *loss* experienced by an individual who chooses an alternative.

In the process of modelling the travel to work, route and departure time selection can be considered as a multiple choice situation in which the utility of selecting option  $i$ , (associated with the choice of a certain route and departure time) is given by :

$$U_i = -w_i = -g.TC_i \quad (2.75)$$

where :

$w_i$  is the perceived disutility of travel associated with selecting option  $i$ , defined as the sum of the measured disutility,  $W_i$  and the random element  $e_i$  :

$$w_i = W_i + e_i \quad (2.76)$$

$TC_i$  is the perceived weighted average of the time loss due to travel time and schedule delay for a traveller selecting option  $i$ .

$g$  is a positive unit scaling between time units and utility units.

Let  $S(w_i)$  denote the expected perceived disutility. A traveller faced with a decision on route and departure time choice will select the alternative that minimises his disutility of travel. Thus his expected perceived disutility of travel is defined as :

$$S(w) = E \left[ \min_{\text{all } j \in C} \{W_j\} \right] \quad (2.77)$$

By substituting eq.(2.75) in eq. (2.77), the latter equation becomes :

$$\begin{aligned} S(w) &= E \left[ \min_{\text{all } j \in C} \{-U_j\} \right] = \\ &= E \left[ \max_{\text{all } j \in C} \{U_j\} \right] = -S(w) \end{aligned} \quad (2.78)$$

The expected perceived disutility function, then, is minus the satisfaction function. The properties of the expected perceived disutility function are similar to those of the satisfaction function except that the former function is related to disutility minimisation while the latter to utility maximisation. Thus :



- i.  $S(w)$  is *concave* with respect to  $w$
- ii. The marginal expected perceived disutility with respect to a certain alternative is the probability of choosing that alternative :

$$\frac{\partial S(w)}{\partial W_i} = P(i) \quad (2.79)$$

- iii.  $S(w)$  is monotonic with respect to the size of the choice set.

## 2.9 Summary

This chapter has presented a framework for the analysis of choice decisions; it was concerned with the conceptual approaches of choice theory and the derivations and properties of some multinomial choice models which will be used in the modelling of dynamic assignment in chapters 6 and 7.

An individual was considered to select a certain alternative following a sequential decision making process where, after defining the choice problem, he generates the set of feasible alternatives, evaluates their attributes, selects an alternative using a decision rule such as dominance, satisfaction or utility and finally implements his decision.

The methodologies described were based on the utility maximisation theory according to which a decision-maker will select the alternative, which for him, has the maximum utility. Utilities are treated as random variables consisting of a measured component which is a function of the attributes of the alternatives and the characteristics of the decision makers, and a random component, due to unobserved or difficult to measure attributes and characteristics, which follows a probability distribution with zero mean.

Different assumptions about the distributions of the random components lead to different choice models. In the case that the random components have a joint normal distribution, choice probabilities are calculated using the multinomial probit model. In terms of realistic choice representation, the MNP is characterised as the most satisfactory form of choice models based on random utility theory, since it allows for any correlation between alternatives. However it does not provide closed form choice probabilities which can only be estimated using approximation or simulation methods. The high computational cost involved when the simulation approach is carried out, is a major disadvantage which makes the MNP unfeasible for applications in large problems; on the other hand the accuracy of the approximation methods is not clear. High computational costs and the questionable accuracy associated with the MNP are the reasons for its limited use in transport demand analysis.

An alternative approach for modelling choice probabilities is the multinomial logit model. The MNL provides the most widely used form of choice models, applied in transportation planning. The main assumption inherent in its formulation is that the utilities of the alternatives are independent and identically distributed Gumbel variables, implying that the ratio of choice probabilities of any two alternatives is not affected by the measured utilities of any other alternatives. This assumption has led to a model of a very simple form which is computationally efficient and therefore appropriate for large scale problems; however it may result in erroneous predictions in the case of correlated alternatives or in the presence of heteroscedasticity (unequal variances) in the utility functions.

The MNL was further extended to the case where the choice set of alternatives is continuous. The new form, termed the continuous logit is useful in the modelling of dynamic assignment since departure time is treated as a continuous variable.

In situations of multidimensional choice it is possible to use a generalisation of the logit model, termed the nested logit, subject to the condition that the correlation between the utilities of the alternatives have a particular structure. The NL avoids the problems associated with the independence of the irrelevant alternatives property of the MNL, since in its formulation the alternatives which are correlated are grouped together and represented by an aggregate alternative. The NL, at the expense of some additional computational requirements, provides a more accurate form than the MNL and will be used in modelling the departure time and route selection.

A dynamic extension of choice models was also presented to estimate the evolution of choice decisions over time. The variable incorporated in dynamic models is the time factor in order to model the individuals' choice decisions which are not static but reviewed over time due to changes in the tastes of individuals and the demand dependent attributes of the alternatives. The models are simplified approximations of a Markov process and provide a useful tool since they do not only calculate the temporal characteristics of choice decisions but also estimate the stationary choice probabilities (as the convergence state when the time factor approaches infinity) in cases where the static models have a complex form which cannot be solved analytically.

Finally the concept of satisfaction and the expected disutility of travel was discussed. The satisfaction of any decision maker is the perceived utility from his choice. Thus in the context of utility maximisation it is defined as the expected value of the maximum perceived utility of the alternatives within the choice set. The expected disutility of travel measures the disutility (due to travel time and schedule delay, in the case of dynamic assignment) that individuals expect to experience when travelling from their

origin to their destination. The properties of the satisfaction function and the expected perceived disutility function were presented and will be used in modelling static and dynamic stochastic assignment in subsequent chapters.

In the following chapter a review of static traffic assignment methods will be presented, with an emphasis on equilibrium assignment algorithms. The choice models described in this chapter will be used in the analysis of stochastic assignment which will form the core of the formulation of dynamic stochastic assignment as an equivalent optimization program in chapter 7.

## **3 static assignment models**

## Objective

The purpose of this chapter is to provide the necessary background information needed for the analysis of traffic assignment and to review some of the various methods of solving the static traffic assignment problem, which will be extended in chapter 7 in order to develop a framework for formulating the dynamic assignment problem as an equivalent optimisation program.

### 3.1 Introduction

Sheffi (1985) defines the problem known as traffic assignment as :

*Given :*

1. *A graph representation of the urban transportation network*
2. *The associated link performance functions*
3. *An origin-destination matrix*

*Find the flow (and travel time) on each of the network links*

Traffic assignment is mainly concerned with the analysis of *road networks*, defined as the physical structures including streets and intersections through which traffic moves. These networks are represented by graphs consisting of nodes and links, the latter being associated with some impedance that affects the flow using it.

The *impedance* or level of service related to the links formulating a network can include many components such as travel time, safety, cost of travel and others. The dominant component is however the travel time since almost all other components are highly correlated with travel time and, it is easier to measure than all the other possible components. Furthermore the level of service provided by many transportation systems

is a function of the demand using these systems. For this reason, in the context of traffic assignment, specific functions were developed to express the relationship between travel time and traffic volume; these functions are referred to as the *link performance or congestion functions*.

The assignment process deals mainly with the allocation of an *origin-destination matrix* onto alternative routes of a network. The O-D matrix, defining the total number of trips for each O/D pair during the period of analysis, is a direct input variable to a traffic assignment model and is either calculated from the previous stages of the transportation planning process (i.e. trip generation, distribution and model split) or can be obtained directly from observed screenline surveys. This chapter focus on the topic of *static network equilibrium* which deals with the assignment of flows in networks assuming that the temporal, or time of day, distribution of the demand is not time dependent (static), i.e. the trip rates between the O-D pairs are assumed to be constant during the period of analysis, and the traffic is considered uniformly distributed within the modelled time interval.

To find the solution to the traffic assignment problem it is necessary to define the rule by which travellers choose a route. It is reasonable to assume that every driver will try to minimise his travel time when travelling from origin to destination. Given the demand between an O-D pair, the question is how the travellers will be distributed among the possible paths connecting this O-D pair. If all of them were to take the shortest path (in terms of travel time), congestion would develop on it. Consequently travel time on this path might increase up to a point where it is no longer the minimum travel time path, and therefore some of the travellers would divert to another path. This alternative path can however be also congested, and so on. The interaction of congestion and travel decisions will finally result in an equilibrium flow pattern. Since, as congestion increases with flow and trips are discouraged by congestion, this interaction can be modelled as a process of reaching an equilibrium between congestion and travel decisions. The concept of equilibrium, in the transportation planning context, was first† introduced, by Wardrop (1952) who stated that traffic on a network distributes itself in such a way that *"the journey times in all the routes actually used are equal, and less than those which would be experienced by a single vehicle on any unused route. ...since it might be assumed that traffic will tend to settle down into*

---

† Beckman *et al.* (1956) points out that the principle of traffic equilibrium was demonstrated (in a network consisting of one high capacity but poor quality road and one lower capacity but better quality road) by Knight F.H. (1924). 'Some Fallacies in the Interpretation of Social Cost'. *Quarterly Journal of Economics*, Vol 38 pp. 582-606.

*an equilibrium situation in which no traveller can reduce his travel time by choosing a new route". This condition is known as the user equilibrium (UE) condition or Wardrop's first principle.*

The user-equilibrium condition includes several assumptions which can be criticised as not realistic. For example travellers are assumed to have full information on the travel times on every route, and to make the correct decisions. However in reality, every traveller may *perceive* a different travel time over the same link, and thus a more realistic decision rule should be that each traveller tries to minimise his perceived travel time. This assumption leads to the *stochastic user equilibrium (SUE)* condition which was introduced by Daganzo and Sheffi (1977) who stated that "*In a stochastic user equilibrium network no user believes he can improve his travel time by unilaterally changing routes*". The stochastic user equilibrium is a generalisation of the user equilibrium definition, since the perceived travel time can be formulated as a random variable distributed across the population of drivers, with mean equal to the actual travel time. Thus if the perception of travel times is accurate, then the SUE flow pattern will be identical to the deterministic UE pattern.

In order for the definition of the above equilibria to be useful, several methodologies have been developed which predict flow patterns that satisfy the UE or the SUE conditions. These are either heuristic methods which attempt to, but do not necessarily provide the equilibrium flows, or they are concerned with the formulation and solution algorithms of the equilibrium assignment problem as an equivalent optimisation program. The equivalent formulation of the network equilibrium was further extended to analyse other stages of the transportation planning process such as trip generation and distribution, and modal split. A review of the alternative formulations and methodologies of the equilibrium approach was done by several researchers including Boyce (1984), Florian (1984), Magnanti (1984), Friesz (1985) and Matsoukis and Michalopoulos (1986). A recent textbook by Sheffi (1985) presents a detailed analysis of the analytical approaches used to study urban transportation network equilibria, and will provide the basis for the analysis that follows in this chapter.

Before proceeding to the presentation of the formulation and solution methodologies of the network equilibrium, the following paragraphs introduce the notation used in the present and following chapters of this thesis. Thus the network is represented by a directed graph defined by a set of consecutively numbered nodes,  $N$ , and a set of consecutively numbered links,  $L$ . A link may also be denoted by its end nodes (i.e., link  $j \rightarrow m$  is the link leading from node  $j$  to node  $m$ ). The set of origin centroids is

denoted by  $R$  and the set of destination centroids by  $S$ . Any origin node may also serve as a destination node, i.e.  $R \cap S \neq \emptyset$ . Each O-D pair  $r$ - $s$  is connected by a set of paths (routes) denoted by  $K_{rs}$ , where  $r \in R$  and  $s \in S$ .

The O-D matrix is denoted by  $Q$  with entries  $Q_{rs}$  representing the trip rate between origin  $r$  and destination  $s$  during the period of analysis. Using vector notation the flow and travel time on a link are denoted by  $q = (\dots, q_a, \dots)$  and  $tt = (\dots, tt_a, \dots)$  respectively, where  $q_a$  represents the flow on link  $a$  and  $tt_a = tt_a(q_a)$  the time needed to traverse link  $a$  when is loaded with flow  $q_a$ . Similarly,  $f_k^{rs}$  and  $c_k^{rs}$  represent the flow and travel time, respectively, on a path  $k \in K_{rs}$  and in vector form are expressed as  $f = (\dots, f^{rs}, \dots)$ ,  $f^{rs} = (\dots, f_k^{rs}, \dots)$  and  $c = (\dots, c^{rs}, \dots)$ ,  $c^{rs} = (\dots, c_k^{rs}, \dots)$ .  $c_k^{rs}$  and  $q_a$  can be expressed mathematically by the following equations, known as the *path-link incidence relationships*.

$$c_k^{rs} = \sum_a tt_a \delta_{a,k}^{rs} \quad \forall k \in K_{rs}, \quad \forall r \in R, \quad \forall s \in S \quad (3.1)$$

$$q_a = \sum_r \sum_s \sum_k \delta_{a,k}^{rs} f_k^{rs} \quad \forall a \in A \quad (3.2)$$

where

$$\delta_{a,k}^{rs} = \begin{cases} 1 & \text{if link } a \text{ is part of path } k \in K_{rs} \\ 0 & \text{otherwise} \end{cases} \quad (3.3)$$

and the corresponding link-path incidence matrix of the O-D pair  $r$ - $s$  is denoted by  $\Delta^{rs}$ .

Having discussed the concept of equilibrium as it is applied in the traffic assignment context, in the next section the network concept is presented, followed by a review on models of link performance functions in section 3.3. With this background material, different approaches to the traffic assignment problem are then presented. Thus section 3.4 describes two heuristic equilibration methods. Section 3.5 presents the mathematical formulation and a solution technique for the equilibrium assignment problem while in section 3.6 the stochastic equilibrium problem is analysed. A summary of the chapter is given in section 3.7.



### 3.2 Network representation

Network is defined as a structure consisting of two types of elements : a set of points and a set of line segments connecting these points. In mathematical terms, network is expressed as a set of nodes and a set of links connecting these nodes. Each link in a network is associated with a direction of flow and usually with some impedance which affects the flow using it. In the networks discussed in this work it is possible to go from any node to any other by following a *path* or *route* through the network. A path is a sequence of directed links leading from one node to another.

Traffic assignment deals exclusively with directed road networks, where all links are directed. The term road network is used to describe the structure containing streets and intersections through which traffic moves. A road network can be represented by a directed graph including nodes and links; links are defined as ordered pairs of nodes including some impedance measures, while nodes are not associated with any impedance measures.

In the transportation planning process urban areas are partitioned into traffic zones, the size of which may vary from a city block to a whole neighbourhood or a suburb. Each traffic zone is represented by a node known as *centroid*. The centroids are those nodes from where traffic originates and to which traffic is destined. Once the set of centroids is defined, the desired movements can be expressed in terms of an Origin-Destination (O-D) matrix. In static assignment this matrix defines the total trip interchange between any O-D pair which is uniformly distributed within the period of analysis. A detailed discussion on different forms of transportation networks and their representation is given by Newell (1980) and others.

The graph representation of a network is not unique but depends on the level of detail at which the network is modelled. Thus for a example the intersection shown in figure 3.1a can be represented as a node (fig. 3.1b) or by a more detailed representation shown in figure 3.1 c. The level of detail depends on the available data and the analysis budget, since the additional links and nodes involved are associated with higher computational costs. On the other hand the simple representation implies that right, left and straight-through movements are equally easy to execute, which is an unrealistic assumption, and also it cannot be used to represent turning restrictions.

The above example of a junction representation leads to the concept of *network aggregation*. The main concern of network aggregation is to create a network smaller than the actual network, either by extracting a subnetwork or by combining detailed

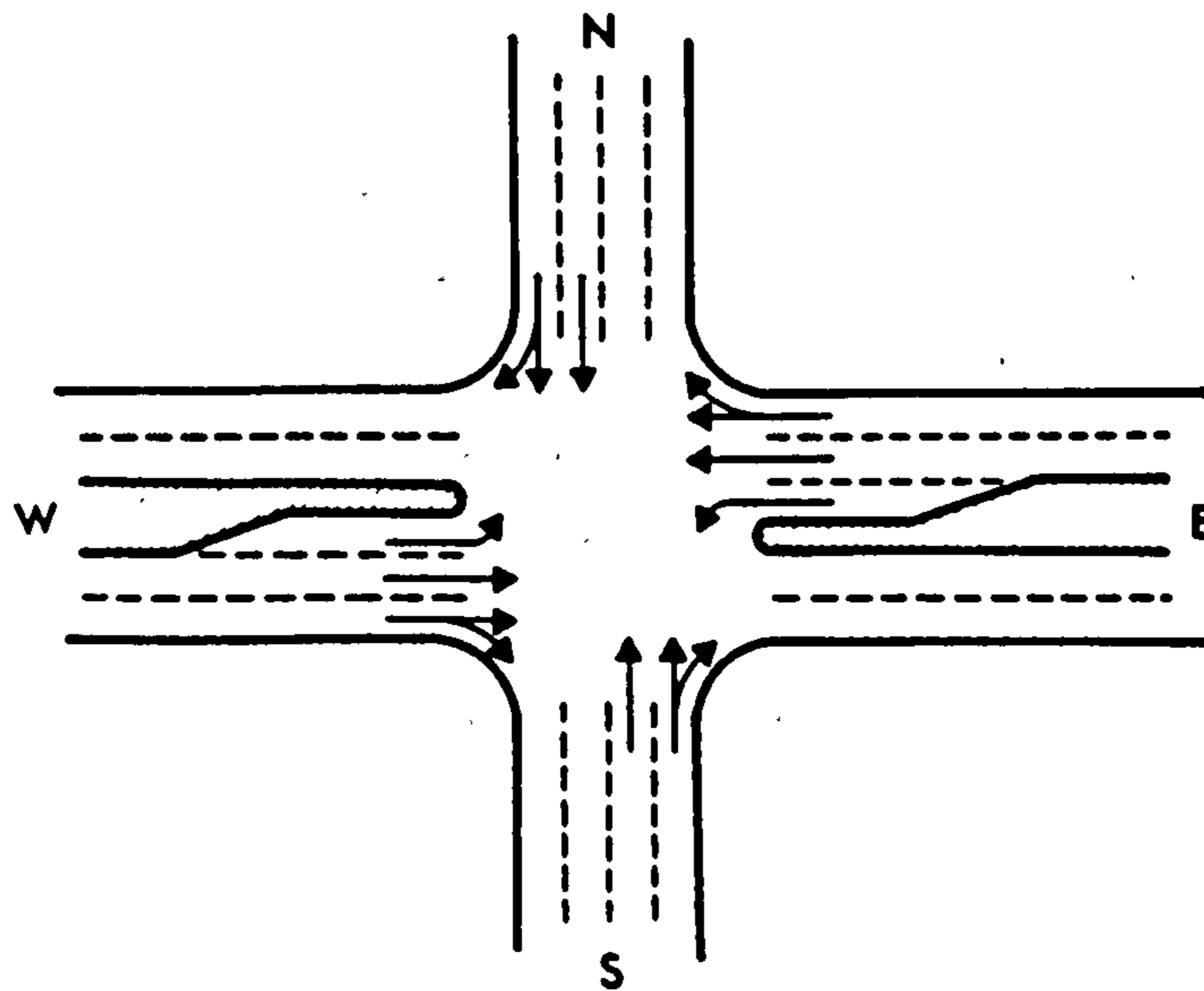


Fig. 3.1a: Layout of a four-approaches intersection.

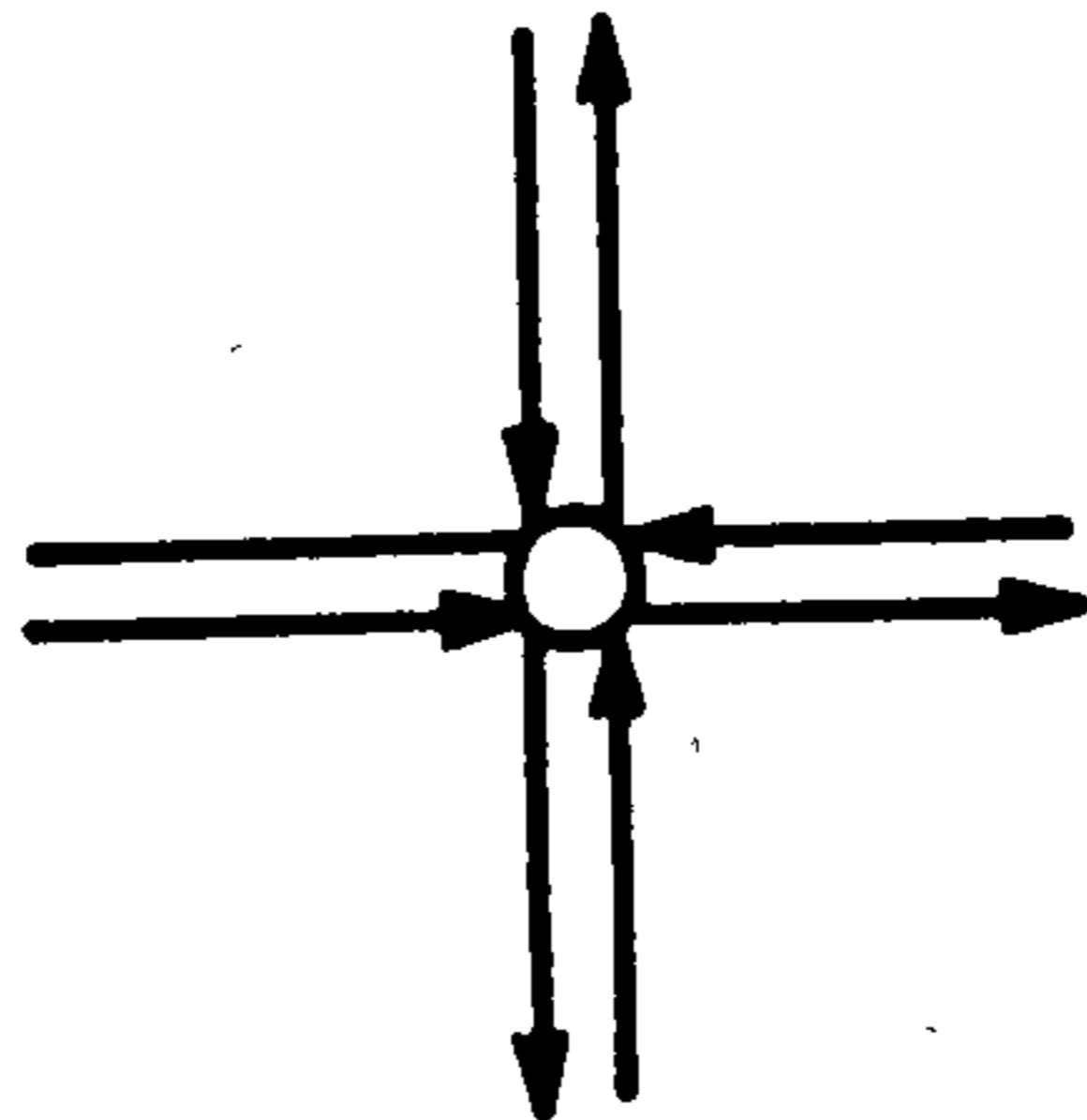


Fig. 3.1b: Network representation of intersection 3.1a as a node.

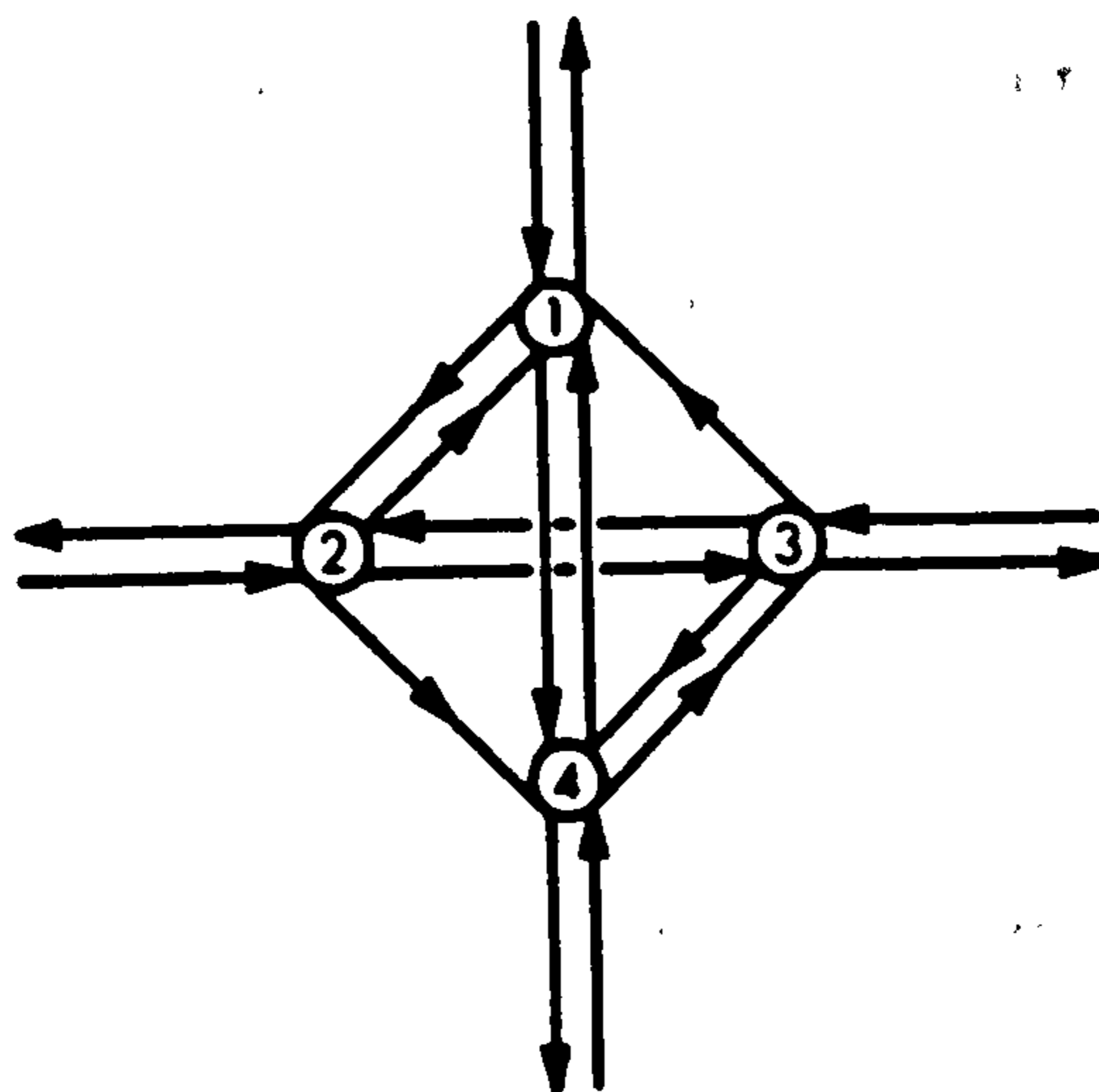


Fig. 3.1c: Detail representation of intersection 3.1a

links into aggregate links, or by both. Network aggregation is an important subject within the traffic assignment framework since the solution of large network equilibrium problems is usually associated with high computational costs.

Several researchers have studied the effect of *spatial detail* (i.e. zone size and network detail) on the accuracy of the resulting estimates of the impacts of a transportation plan. Among them, Bovy and Jensen (1983) used the network of Eindhoven, the Netherlands, and Eash *et al.* (1983) used the network of the Chicago regional area to deal with the problem of how different network representations affect the final results of equilibrium assignment. They concluded that *ad hoc* aggregation procedures often produce unreliable flow patterns particularly when the network of interest is significantly congested. The more detail the representation of the network, the better the final results are likely to be; but there is a limitation by the computational cost caused by additional links, nodes and centroids.

Most methods require that all possible origins and destinations of trips taking place within an area be represented as if they were taking place to and from a small set of points or centroids. In order to overcome this problem, which is known as the *spatial aggregation* problem one could use smaller zones and more centroids, but this would imply a high increase in the computational effort. Daganzo (1980a,b) introduces an algorithmic procedure which is designed to handle a substantially larger number of centroids, and also to take account of a continuous distribution of population.

When there is no congestion, as Zipkin (1980) argues, the network aggregation problem is not difficult. However in the case of congestion the problem becomes rather complex and mainly deals with the extraction of a subnetwork, the approximation of the flows entering and exiting the subnetwork, and the construction (using these flows) of an O-D matrix which will allow a user equilibrium flow pattern to be calculated. Algorithms for this type of aggregation problems has been proposed by Haghani and Daskin (1983) and Hearn (1976) who also reviews different network aggregation techniques.

### 3.3 Link performance functions

The level of service provided by many transportation systems is a function of the demand using these systems. Because of congestion, travel time is an increasing function of flow. Therefore a link performance function (rather than a constant travel

time value) relating the travel time needed to traverse a link, to the flow traversing this link, should be associated with each of the links representing an urban network.

Branston (1976) reviewed several link performance functions in relation to their ability to represent observed data and their applicability in different assignment procedures and argued that *"there seems to be very little agreement between researchers on the type of function which is suitable for any particular network"*. In this section, some of the most commonly used functions are presented.

One of the best known and most widely used congestion function to have been developed is the one often referred to as the BPR function, Bureau of Public Roads (1964), or alternatively FHWA (Federal Highway Administration) function. In its general form it is expressed as :

$$tt(q) = tt_0 \cdot [1 + a(q/c)^b] \quad (3.4)$$

where :

- q            the assigned flow
- tt(q)       the average travel time
- tt<sub>0</sub>        the free flow travel time
- c            the practical capacity †.
- a,b         positive constants. Values suggested in BPR(1964) were 0.15 and 4, respectively.

The main limitations of this formula is the lack of a theoretical foundation and that the definition of the capacity in the function is rather ambiguous.

The above formula has been used in many transportation planning studies and incorporated in several network design models, including the ones by Abdulaal and LeBlank (1979), LeBlank (1975), Poorzahedy and Turnquist(1982) and others. Furthermore as was mentioned by Gartner *et al.* (1980) it *"..has been (and continues to be) the single most important analytical model used in the urban transportation planning process. Yet the origins of this function, or the experimental data on which it is based, are obscure..."*.

---

† Practical capacity is defined in the Highway Capacity Manual (BRP 1950) as *"the maximum number of vehicles that can pass a given point on a roadway or in a designated lane during one hour without the traffic density being so great as to cause unreasonable delay, hazard, or restriction to the drivers' freedom to manoeuvre under the prevailing roadway and traffic conditions"*. The values of a=0.15 and b=4 imply that the practical capacity is the flow at which travel time is 15% higher than the free flow travel time.

Davidson (1966) has suggested a different congestion function which relates free-flow travel time to saturation flow and delay, and is expressed as :

$$tt(q) = tt_0 \cdot [ 1 + Jq/(s-q) ] \quad (3.5)$$

where  $s$  is the saturation flow and  $J$  a delay parameter

The Davidson function has the ability to include the effects of road type and environment through the parameters  $s$  and  $tt_0^\dagger$ , and is supported by a theoretical justification using queueing theory (Davidson (1978)).

However its use may lead to computational difficulties when an iterative assignment procedure is applied (as the Frank and Wolfe algorithm which is analysed in section 3.5), if oversaturated conditions occur, which often do at any stage during the execution even if the final flow conditions are not oversaturated. Daganzo (1977a,b) has solved this problem by limiting link volumes to capacity, However as Taylor (1984) argues, *the restriction of link flow levels needs to be considered carefully*, since in studies of network flows over restricted time intervals (e.g. peak period), the typical situation to which equilibrium assignment is applied, *the role of a link as an input-storage device in a network becomes important*. Although the output of a bottleneck cannot exceed its capacity, the input, i.e. the demand to use the bottleneck, may, and the result is the formation of queues. Ignoring this process may result in unrealistic predictions. *This is an implicit consequence of the use of a steady state procedure to model a dynamic situation.*

Boyce *et al.*(1981) evaluated the performance of the Davidson and the FHWA function in equilibrium assignment by comparing the resulting assignments to observed traffic counts from the Chicago Metropolitan area. They showed that Davidson's functions may lead to unrealistic high travel times, since travel time estimations tend to infinity as link flow approaches the saturation flow, affecting both individual link flows, and the total system vehicle km.

Akcelik (1978) has suggested a continuous version of the Davidson function yielding finite travel times for all finite flows which allows short term overloading of links in an assignment and thus avoids the computational difficulties mentioned by Boyce *et al* (1981). Matsoukis and Michalopoulos (1986) argue that the modified Davidson function is " *useful for inclusion in an equilibrium assignment given its ability to reflect differences in network link type (capacity and speed) and environment through its*

---

† Taylor (1977) has developed a method for the direct estimation of these parameters using the method of least squares.

parameters, and the conceptual advantage of the function through its derivation from queueing theory".

This function is expressed as :

$$tt(q) = tt_0 [ 1 + Jq/(s-q) ] \quad q \leq \tau s \quad (3.6a)$$

$$= tt_\tau + K_\tau(q-\tau s) \quad q > \tau s \quad (3.6b)$$

where

$$0 < \tau < 1 \quad (3.7)$$

$$tt_\tau = tt(\tau s) = tt_0 [ 1 + J\tau/(1-\tau) ] \quad (3.8)$$

$$K_\tau = \left[ \frac{dtt}{dq} \right]_{q=\tau s} = Jtt_0/(s(1-\tau)^2) \quad (3.9)$$

The critical flow  $\tau s$  may be considered as the flow level above which oversaturation of junction occurs and at which junction delays consequently dominate link travel times. The factor  $\tau$  becomes the fourth parameter of this model. Ackelik (1978) suggested that values of  $\tau$  in the range (0.85 - 0.90) could be selected to reflect the quality of service provided by a particular road.

Taylor (1984) used the modification of the Davidson's and the FHWA congestion function in the equilibrium assignment of the Melbourne network. He found that the assignments produced a satisfactory degree of fit to the observed link flows, given the biases and possible errors present in the available data and also that there was little difference between the assigned volumes based on the two different congestion functions.

### 3.4 Heuristic equilibration techniques

An essential procedure used in the heuristic equilibration as well as in the UE methods is the *network loading* mechanism. Network loading is the process of assigning demand to the network in which link travel times are considered constant. This process is executed assuming that the individuals travelling between an O-D pair select the route that is associated with the shortest travel time, and is known as the *all-or-nothing* assignment, since all the other paths connecting this O-D pair do not carry any flow.

In the all-or-nothing procedure, each O-D pair is examined in turn and the trip interchange is assigned to every link that is on the minimum time path connecting the O-D pair. The assigned volumes are accumulated for each link and the total volume is the sum of all individual link volumes for each trip interchange. In this procedure the dependence of travel time on traffic flow is not considered, and thus the equilibrium concept is ignored. It was used in the early urban transportation studies as the traffic assignment procedure.

### Capacity Restraint

The initial form of this method involves a repetitive all-or-nothing assignment in which the travel times from the previous assignment are used in the current iteration. This process of iteration is continued until the resulting flows from the previous iteration are very similar to the ones derived from the current iteration. However this procedure may result in an situation where in every iteration the O-D demand interchanges between two paths, without loading any other paths of the network and consequently not converging to any solution. To avoid this situation, instead of using the travel time resulting from the previous iteration for the new loading, a combination of the last two travel times obtained is used. This introduces a *smoothing* effect, but does not guarantee a convergence to an equilibrium. The algorithm terminates after a predetermined number of iterations, and summarised as follows :

#### Step 0: *Initialisation.*

Perform an all-or-nothing assignment based on  $tt_a^0 = tt_a(0)$ ,  $\forall a$ . Obtain a set of link flows  $\{q_a^0\}$ . Set iteration counter  $n = 1$ .

#### Step 1: *Update.*

Set  $\tau_a^n = tt_a(q_a^{n-1})$ ,  $\forall a$

#### Step 2: *Smoothing.*

Set  $tt_a^n = 0.75tt_a^{n-1} + 0.25\tau_a^n$ ,  $\forall a$ .

#### Step 3: *Network loading.*

Perform all-or-nothing assignment based on travel times  $\{tt_a^n\}$ . This yields a set of link flows  $\{q_a^n\}$ .

#### Step 4: *Stopping rule.*

If  $n = N$ , go to step 5. Otherwise, set  $n = n + 1$  and go to step 1.

**Step 5: Averaging.**

Set  $q_a^* = (1/4) \sum_{i=0}^3 q_a^{n-i}$ ,  $\forall a$  and stop. The set  $\{q_a^*\}$  provides the link flows at equilibrium.

**Incremental Assignment**

Another heuristic equilibration technique is the incremental assignment; in this procedure a portion of the O-D matrix is assigned at each iteration. The travel times are then updated and an additional portion of the O-D matrix is loaded onto the network. The algorithm can be summarized as follows :

**Step 0: Preliminaries.**

Divide each origin-destination entry into  $N$  equal portions (i.e. set  $Q_{rs}^n = Q_{rs}/N$ ). Set iteration counter  $n = 1$  and  $q_a^0 = 0$ ,  $\forall a$ .

**Step 1: Update.**

Set  $tt_a^n = tt_a(q_a^{n-1})$ ,  $\forall a$

**Step 2: Incremental loading.**

Perform all-or-nothing assignment based on travel times  $\{tt_a^n\}$ , but using only the trip rates  $Q_{rs}^n$ . This yields a set of link flows  $\{x_a^n\}$ .

**Step 3: Flow summation.**

Set  $q_a^n = q_a^{n-1} + x_a^n$ ,  $\forall a$ .

**Step 4: Stopping rule.**

If  $n = N$ , stop (the current set of link flows is the solution). Otherwise, set  $n = n + 1$  and go to step 1.

The incremental assignment technique does not guarantee a convergence to an equilibrium flow pattern, and since it terminates after a predetermined number of iterations it may converge to a non equilibrium solution.

A review on the heuristic equilibration techniques and other non-equilibrium methods for solving the traffic assignment problem is presented in a recent paper by Matsoukis (1986).



### 3.5 User equilibrium assignment

The heuristic methods presented in the previous section do not necessarily converge to an equilibrium solution. The equilibrium assignment problem is to find the link flows,  $q$ , that satisfy the Wardrop's equilibrium principle. In this section the formulation and solution of the UE assignment problem as an equivalent minimisation program is presented.

#### Program Formulation

Beckman *et al.*(1956) were the first to formulate the equilibrium assignment problem with elastic demand, as a minimisation problem with linear constraints; they also proved the equivalency, existence and uniqueness of the solution of the minimisation program subject to the condition that the link performance function on each link is an increasing function of the flow on the link. The simplest form of this program, known as the *Beckman's transformation* (Sheffi (1985)), is for the case that the demand is fixed, and is expressed as :

$$\min_{\mathbf{q}} z(\mathbf{q}) = \sum_a \int_0^{q_a} tt_a(\omega) d\omega \quad (3.10a)$$

subject to

$$\sum_k f_k^{rs} = Q_{rs} \quad \forall r,s \quad \text{and} \quad f_k^{rs} \geq 0 \quad \forall k,r,s \quad (3.10b)$$

The definitional constraints :

$$q_a = \sum_r \sum_s \sum_k f_k^{rs} \delta_{a,k}^{rs} \quad \forall a \quad (3.10c)$$

are also part of this program.

#### Equivalency Conditions

In order to show that program (3.10) is equivalent to the equilibrium assignment problem it has to be proved that any flow pattern that solves program (3.10) also satisfies the UE conditions. At the solution point the first order conditions must be satisfied. However, the first order condition of program (3.10) are equivalent to the first order conditions of the Lagrangian :

$$L(\mathbf{f}, \mathbf{u}) = z[\mathbf{q}(\mathbf{f})] + \sum_{rs} u_{rs} ( Q_{rs} - \sum_k f_k^{rs} ) \quad (3.11)$$

given that  $L(\mathbf{f}, \mathbf{u})$  has to be minimised with respect to nonnegative path flows, i.e :

$$f_k^{rs} \geq 0 \quad \forall k,r,s \quad (3.12)$$

where  $u_{rs}$  denotes the dual variable associated with the flow conservation constraint for the O/D pair  $r/s$ .

The formulation of the Lagrangian is given in terms of path flow ( $q_a = q_a(f)$ ) using the incidence relationship (3.10c). At the minimum point of the Lagrangian, the following conditions must hold with respect to the path flow variables :

$$f_k^{rs} \frac{\partial L(f, u)}{\partial f_k^{rs}} = 0 \quad \forall k, r, s \quad \text{and} \quad \frac{\partial L(f, u)}{\partial f_k^{rs}} \geq 0 \quad \forall k, r, s \quad (3.13)$$

and the following conditions must hold with respect to the dual variables :

$$\frac{\partial L(f, u)}{\partial u_{rs}} = 0 \quad \forall r, s \quad (3.14)$$

Also the nonnegativity constraints (3.12) have to hold. The above conditions are proved to be equivalent to :

$$f_k^{rs} (c_k^{rs} - u_{rs}) = 0 \quad \forall k, r, s \quad (3.15)$$

$$c_k^{rs} - u_{rs} \geq 0 \quad \forall k, r, s \quad (3.16)$$

$$\sum_k f_k^{rs} = Q_{rs} \quad \forall r, s \quad (3.17)$$

$$f_k^{rs} \geq 0 \quad \forall k, r, s \quad (3.18)$$

Conditions (3.17) and (3.18) are the flow conservation and the nonnegativity constraints. Conditions (3.15) and (3.16) hold for each path between any O/D pair in the network. Thus for a path  $k$  connecting origin  $r$  to destination  $s$ , the conditions hold for two possible combinations of path flow and travel time :

- i) Either the flow on path is zero, i.e.  $f_k^{rs} = 0$  and eq.(3.15) holds, in which case the travel time on this path,  $c_k^{rs}$ , must be greater than or equal to the O/D Lagrange multiplier  $u_{rs}$ , as required by condition (3.16), or
- ii) the flow on the  $k$  path is positive, in which case  $c_k^{rs} = u_{rs}$ , and both conditions (3.15) and (3.16) hold as equations.

Equations (3.15) and (3.16) show that the Lagrange multiplier  $u_{rs}$  of a given O-D pair is less than or equal to the travel time on any path connecting this pair, implying that  $u_{rs}$  equals the minimum path travel time between origin  $r$  and destination  $s$ . Therefore any path that carries flow is associated with a travel time equal to the minimum O-D travel time. If the flow pattern satisfies these equations, then no traveller can reduce his travel time by choosing a new route.

### Uniqueness Conditions

The uniqueness of the optimal solution of the minimisation program (3.10) is ensured if the objective function  $z(q)$  is strictly convex in the vicinity of the optimum solution (and convex elsewhere). This is done by proving that the Hessian (the matrix of the second derivatives of  $z(q)$  with respect to  $q$ ) is positive definite, ensuring that  $z(q)$  is strictly convex everywhere. However the Hessian is a diagonal matrix with positive entries, since the link performance functions considered are increasing functions of the flow on the link, and therefore it is a positive definite matrix.

### Solution algorithm

Dafermos and Sparrow (1969) have proposed a formulation similar to the one proposed by Beckman *et al.* and were the first to provide a solution to the UE assignment problem; however the algorithm they developed has not been adopted for the large scale networks considered in planning practice. Another approach was suggested by Nguyen (1974); his solution algorithm uses the Convex-Simplex Method (Zangwill (1969)), and was adopted by Florian and Nguyen (1976) in the study of the network of the City of Winnipeg providing convincing results.

However the most widely used technique to solve the UE problem is the convex combinations method, developed by Frank and Wolfe (1956) and therefore known as the Frank and Wolfe (F-W) algorithm. This algorithm was first applied to UE assignment by Murchland (1969), and later by Leblanc *et al.* (1975) in a test network including 76 links and 24 nodes.

The F-W algorithm is an iterative procedure which is based on choosing a descent direction at the current solution point  $q^n$  and moving along this direction to the next solution point,  $q^{n+1}$ , so that  $z(q^{n+1}) < z(q^n)$ . This algorithmic step can be written in as :

$$q^{n+1} = q^n + \alpha_n d^n \quad (3.19)$$

where  $d^n$  is a descent direction vector and

$\alpha_n$  is a nonnegative scalar known as the *move size*

The descent direction is defined as  $d^n = y^n - q^n$ , where  $y^n$  is defined from the following minimisation program :

$$\min z^n(y^n) = \nabla z(q^n) \cdot y^{nT} = \sum_a \frac{\partial z(q^n)}{\partial q_a} y_a = \sum_a t t_a^n y_a^n \quad (3.20)$$

subject to

$$\sum_k f_k^{rs} = Q_{rs} \quad \forall r,s \quad \text{and} \quad f_k^{rs} \geq 0 \quad \forall k,r,s \quad (3.21)$$

where  $q_a = \sum_k \sum_{rs} f_k^{rs} \delta_{a,k}^{rs}$ ,  $\forall a$ , and  $tt_a = tt_a(q_a^n)$ , and  $y_a$  is the auxiliary variable representing the flow on link  $a$ . The above program is obviously minimised when  $y^n$  are the flows resulting from an all-or-nothing assignment.

Furthermore the new solution  $q^{n+1}$ , must lie between  $q^n$  and  $y^n$ , since  $y^n$  lies on the boundary of the feasible region defined by the constraints. Thus, the move size is determined by minimising the objective function along the descent direction and defined as :

$$\min_{0 < \alpha < 1} z[q^n + \alpha(y^n - q^n)] \quad (3.22)$$

A possible criterion to indicate the convergence of the algorithm towards an equilibrium solution is based on the similarity of successive O-D travel times or link flows.

The algorithm as applied in equilibrium assignment is summarised as follows :

**Step 0 : Initialisation.**

Perform an all-or-nothing assignment based on  $tt_a^0 = tt_a(0)$ ,  $\forall a$ . Obtain a set of link flows  $\{q_a^1\}$ . Set iteration counter  $n = 1$ .

**Step 1: Update.**

Set  $t_a^n = tt_a(q_a^n)$ ,  $\forall a$

**Step 2: Direction finding.**

Perform an all-or-nothing assignment based on  $\{t_a^n\}$ . This yields a set of auxiliary link flows  $\{y_a^n\}$ .

**Step 3: Line Search.**

Find  $\alpha_n$  that solves

$$\min_{0 \leq \alpha \leq 1} \sum_a \int_0^{q_a^n + \alpha(y_a^n - q_a^n)} tt_a(\omega) d\omega$$

**Step 4: Move.**

Set  $q_a^{n+1} = q_a^n + \alpha_n(y_a^n - q_a^n) \quad \forall a$

**Step 5: Convergence test.**

If a convergence criterion is met, stop (the current solution,  $\{q_a^{n+1}\}$ , is the set of equilibrium link flows); otherwise, set  $n=n+1$  and go to step 1.

Eash et al (1979) implemented the first convergent version of the F-W to be used regularly in transportation planning practice in the Chicago Area Transportation Study. The F-W algorithm was further implemented in several transportation planning studies.

Dow (1979) has shown that the speed of convergence may be improved by using a good initial solution. Additional research by Fukushima (1984), Weintraub *et al.* (1985), Lupi (1986) and others has led to some improvements in the speed of convergence of the F-W.

**Extension of User Equilibrium**

The framework of the equilibrium analysis was extended to include cases in which travel time over a link depends on the flow on the link as well as on the flow on other links, as is the case of two-way streets, intersections, merging movements and turning movements. The concept of link interactions was introduced by Dafermos (1971) and was analysed by several researchers including Smith(1979a), Dafermos (1980,1982), Fisk and Nguyen (1982), Heydecker (1983), Nguyen and Dupuis (1984) and others.

The UE formulation was also extended and generalised to the joint treatment of several travel-choice dimensions, such as whether to travel or not, where to go, which mode to use and what route to take. Thus, Evans (1976) has developed a model to further include trip distribution, while Abdulaal and Leblanc (1979) studied the problem of combined modal split and traffic assignment. A combined model of trip mode choice, trip distribution and traffic assignment was studied by Florian and Nguyen (1978) who modelled the "four stages" transportation modelling process by adopting an equilibrium formulation for the simultaneous choice of destination, mode and route, instead of considering transportation modelling as a sequential process.

A survey on the recent developments on user equilibrium as well as on stochastic user equilibrium, which will be analysed in the following section, is given by Florian (1984), Magnanti (1984) and Boyce (1984) and Friesz (1985) who also describes the basic research which must be conducted to advance these fields.

### 3.6 Stochastic Equilibrium

A main assumption in the deterministic approach to traffic assignment is that travellers have perfect information on the link travel times over the entire network, and make the correct decisions with regards to the choice of the shortest travel time route. Stochastic assignment relaxes some of these assumptions by modelling link travel times as random variables, and analysing route selection using the concepts of utility maximisation and random utility.

At the core of the stochastic equilibrium assignment methods lies the *stochastic network loading* (SNL) mechanism. Following the definition of network loading given in section 3.4, SNL is the process of assigning demand to a network in which link travel times are assumed random, but not flow dependent, in contrast to the SUE assignment where link travel times are assumed random but also flow dependent.

This section will introduce some of the concepts of stochastic assignment, with emphasis on logit-based formulations. Thus in subsection 3.6.1 some SNL algorithms will be presented, while in 3.6.2 the formulation and solution of the SUE assignment problem as an equivalent minimisation program is presented.

#### 3.6.1 Stochastic network loading

The SNL models are a special case of multiple choice models. The flow pattern resulting from a SNL is defined by first calculating the path choice probabilities and then implementing the link-path incidence relationship (eq. (3.10c)). The most widely used SNL algorithms are the ones suggested by Burrell (1968), Dial (1971) and Daganzo and Sheffi (1977).

In Burrell's (1968) method each origin is examined in turn. For each origin a set of perceived link travel times is obtained, by sampling each link travel time from a discrete uniform distribution with variance to mean ratio equal on all links. The corresponding tree is then defined, and all the trips emanating from that origin are then loaded on the network, by assigning the trip interchange between each O-D pair to every link that is on the path which is characterised by the minimum perceived travel time.

However the most popular SNL model is the Dial's (1971) method, also known as the **STOCH** algorithm. This algorithm will be presented here in more detail.

The **STOCH** algorithm is a procedure that allocates travel demands among alternative paths by means of a logit model. However, it does not assign choice probabilities to all the paths connecting each O/D pair. Instead it assumes that many of these paths constitute unreasonable travel choices that would not be considered in practice. A path is reasonable if every link in it has its initial node closer to the origin than its final node, and has its final node closer to the destination than its initial node. A more efficient algorithm can be developed by redefining the criteria for reasonable paths. Thus a *path would now be considered reasonable if it includes only links that do not take the traveller back towards the origin.* The steps of the **STOCH** algorithm are described below :

**Step 0: Preliminaries**

To assign  $x$  trips between origin node  $r$  to a destination node  $s$ , the following four variables must be computed for each node  $i$  :

- a)  $r(i)$  = the shortest path distance from  $r$  to  $i$
- b)  $I_i$  = the set of downstream nodes of all links leaving node  $i$ .
- c)  $F_i$  = the set of upstream nodes of all links arriving at node  $i$ .
- d)  $L(i \rightarrow j)$  = the *link likelihood* of link  $i \rightarrow j$  expressed as :

$$L(i \rightarrow j) = \begin{cases} e^{\mu[r(j)-r(i)-t(i \rightarrow j)]} & \text{if } r(i) < r(j) \\ 0 & \text{otherwise} \end{cases} \quad (3.23)$$

In this expression  $t(i \rightarrow j)$  is the measured travel time on link  $i \rightarrow j$ .

**Step 1: Forward pass**

Examine nodes in ascending order of  $r(i)$  starting from the origin,  $r$ . For each node,  $i$ , calculate the *link weight*  $w(i \rightarrow j)$  for each node  $j \in I_i$

$$w(i \rightarrow j) = \begin{cases} L(i \rightarrow j) & \text{if } i = r \text{ (i.e., if node } i \text{ is the origin)} \\ L(i \rightarrow j) \sum_{m \in F_i} w(m \rightarrow i) & \text{otherwise} \end{cases} \quad (3.24)$$

When the destination node,  $s$ , is reached, this step is complete.

**Step 2: Backward pass**

Starting with the destination,  $s$ , examine all nodes,  $j$ , in ascending order of  $s(j)$ . For each node,  $j$ , calculate the link flow  $q(i \rightarrow j)$  for each  $i \in F_j$  using the formula :

$$q(i \rightarrow j) = \left[ q_{rj} + \sum_{m \in I_i} q(j \rightarrow m) \right] \frac{w(i \rightarrow j)}{\sum_{m \in F_i} w(m \rightarrow j)} \quad (3.25)$$

Where  $q_{rj}$  is the trip rate from the origin node,  $r$ , to node  $j$ , and  $q(i \rightarrow j)$  is defined as zero if  $I_j$  is empty (i.e. for destinations on the edge of the network).

The steps terminates when the origin node,  $r$ , is reached.

Daganzo and Sheffi (1977) propose a different approach which applies the multinomial probit in order to calculate the choice probability of a route between an origin and a destination. The only feasible method to calculate the choice probabilities is suggested by Sheffi and Powell (1981) and is based on Monte Carlo simulation. The algorithm is an iterative procedure, where at each iteration a set of perceived link travel times is obtained, by sampling each link travel time from a normal distribution. This set of perceived travel times is then used in an all-or-nothing assignment and the demand between each O-D pair is assigned to the shortest travel time path based on the simulated perceived travel times. The process of sampling and assignment is repeated several times, and when the process terminates the results of the individual iterations are averaged for each link to give the final flow pattern.

The probit-based SNL mechanism is based on realistic assumptions but is associated with high computational effort, in contrast to the STOCH algorithm which is very computationally efficient. However as Burrell (1976), Florian and Fox (1976), Daganzo and Sheffi (1977) and others pointed out, the latter method tends to overload overlapping paths. This is because the probability of choosing a path is assumed to be independent from its overlap with other paths. This deficiency is caused by the independence of irrelevant alternatives property of the logit model. Tobin (1977) has extended Dial's algorithm to overcome the limitations of the logit formulation. However Sheffi (1979) argues that this new development is associated with other deficiencies which do not exist in Dial's original procedure and should be subject to further research before implementation.



### 3.6.2 Stochastic user equilibrium assignment

In the stochastic equilibrium assignment the perceived travel times are modelled not only as random variables but also as flow dependent, by assuming that the mean travel time for each link is a function of flow on that link. Therefore  $tt_a = tt_a(q_a)$  and also  $tt_a = E[T_a]$ , where  $T_a$  is the perceived travel time on a link  $a$ .

Given the O-D trip rates,  $\{Q_{rs}\}$ , the stochastic equilibrium conditions can be characterised by the following equations :

$$f_k^{rs} = Q_{rs} P_k^{rs} \quad (3.26)$$

implying that :

$$q_a = \sum \sum Q_{rs} P_k^{rs} \delta_{a,k}^{rs} \quad \forall a \quad (3.27)$$

where

$$P_k^{rs} = P_k^{rs}(tt) = \Pr(C_k^{rs} \leq C_m^{rs}, \forall m \neq k \in K_{rs} | tt) \quad (3.28)$$

and expressing the probability that a driver travelling between the O-D pair  $r$ - $s$  will select route  $k \in K_{rs}$

$C_k^{rs}$  is the random variable representing the perceived travel time on route  $k$  between  $r$  and  $s$ , defined as:

$$C_k^{rs} = \sum_a T_a \delta_{a,k}^{rs}, \forall k, r, s. \quad (3.28)$$

#### Program Formulation

Fisk (1980) was the first to give a solution to the SUE assignment problem. She has formulated a network optimisation problem whose solution results in a traffic assignment in which path choice is governed by a logit distribution incorporating flow dependent path costs. She also showed that in SUE assignment the undesirable characteristics of the logit formulation (inherent in Dial's model) appear when the logit scale parameter  $\mu \rightarrow 0$ . However she argues that as  $\mu$  increases the effects of the IIA property become less predominant, and the magnitude of these effects is not important for realistic values of  $\mu$ .

The formulation of SUE as an equivalent minimisation program which is presented below was suggested by Sheffi and Powell (1982) who also proved the uniqueness of its solution. Their work extends the earlier work by Daganzo (1979) in the context of discrete choice models and is analysed here in more detail, since it provides the basis of the dynamic stochastic user equilibrium formulation as an equivalent optimisation

program, which will be presented in chapter 7. The SUE equivalent mathematical program is expressed as :

$$\min_{\mathbf{q}} z(\mathbf{q}) = - \sum_{rs} Q_{rs} E \left[ \min_{k \in K_{rs}} \{C_k^{rs}\} | c^{rs}(\mathbf{q}) \right] + \sum_a q_a tt_a(q_a) - \sum_a \int_0^{q_a} tt_a(\omega) d\omega \quad (3.29)$$

which is equivalent to :

$$\min_{\mathbf{q}} z(\mathbf{q}) = - \sum_{rs} Q_{rs} S_{rs}[c^{rs}(\mathbf{q})] + \sum_a q_a tt_a(q_a) - \sum_a \int_0^{q_a} tt_a(\omega) d\omega \quad (3.30)$$

since as was defined in eq(2.78)

$$S_{rs}[c^{rs}(\mathbf{q})] = E \left[ \min_{k \in K_{rs}} \{C_k^{rs}\} | c^{rs}(\mathbf{q}) \right] \quad (3.31)$$

### Equivalency Conditions

In order to show that program (3.29) is equivalent to the SUE problem it has to be proved that any flow pattern that solves program (3.29) also satisfies the SUE conditions. The first-order conditions for unconstrained minimisations require only that the gradient vanishes at the minimum point. Taking into account that † :

$$\frac{\partial S_{rs}(c^{rs})}{\partial c^{rs}} = P_k^{rs} \quad (3.32)$$

The first order conditions are expressed as :

$$\nabla z(\mathbf{q}) = \left[ - \sum_{rs} Q_{rs} P_k^{rs} \Delta^{rs T} + \mathbf{q} \right] \cdot \nabla_{\mathbf{q}} tt = 0 \quad (3.33)$$

where a typical term of the above expression is :

$$\frac{\partial z(\mathbf{q})}{\partial q_a} = \left[ - \sum_{rs} \sum_k Q_{rs} P_k^{rs} \delta_{a,k}^{rs} + q_a \right] \cdot \frac{dtt_a}{dq_a} = 0 \quad \forall a \quad (3.34)$$

Assuming that the link performance functions are strictly increasing (i.e.  $dtt_a/dq_a > 0$ ), the gradient of  $z(\mathbf{q})$  can only vanish only if

$$q_a = \sum_{rs} \sum_k Q_{rs} P_k^{rs} \delta_{a,k}^{rs} \quad \forall a \quad (3.35)$$

The above condition is identical to the SUE condition expressed in eq. (3.27).

---

† From equation (2.79)

### Uniqueness Conditions

In order to demonstrate the uniqueness of the optimum solution the Hessian matrix of the objective function must be proved to be a positive definite matrix, implying that the program is strictly convex. However this matrix, is not in general positive definite, except of the area where the equilibrium point is approached, i.e when  $(q_a - \sum_{rs} \sum_k Q_{rs} P_k^{rs} \delta_{a,k}^{rs}) \rightarrow 0$ , implying that the equilibrium point is in fact a local minimum, but not ensuring that it is the only local minimum. However the objective function (3.29) can be expressed as function of  $tt$ , by introducing the relationship  $q_a = q_a(tt_a)$ :

$$z(tt) = - \sum_{rs} Q_{rs} E \left[ \min_k (C_k^{rs} | c^{rs}) \right] + \sum_a q_a(tt_a) tt_a - \sum_a \int_{tt_a(0)}^{tt_a} \omega \frac{dq_a(\omega)}{d\omega} d\omega \quad (3.36)$$

whose gradient has a typical term of the form :

$$\frac{\partial z(q)}{\partial x_a} = - \sum_{rs} \sum_k Q_{rs} P_k^{rs} \delta_{a,k}^{rs} + q_a \quad \forall a \quad (3.37)$$

The Hessian of the new expression of the objective function is a positive definite matrix and thus  $z(tt)$  is strictly convex, having a single stationary point which is its minimum. However the functions  $z(q)$  and  $z(tt)$  are related by a monotonic transformation, and the gradients of both functions always have the same sign and vanish at the same points. Furthermore, the gradient of  $z(tt)$ , vanishes only once, at the minimum of  $z(tt)$ ; therefore  $z(q)$  must also have a unique minimum at this point.

### The solution algorithm

The most popular algorithm used to solve the equivalent minimisation program is the method of successive averages (MSA), (described by Wilde (1964)). This algorithm was first used in traffic assignment by Sheffi and Powell (1981) to solve heuristically the SUE problem and was suggested by the same authors (1982) as the solution algorithm of the minimisation program (3.29) who also proved its convergence (Powell and Sheffi (1982)) when applied in network equilibrium problems.

The MSA is a descent direction algorithm where the move size  $\alpha_n$  is not determined from some characteristics of the current solution, but it is a predetermined sequence which has to satisfy the following two conditions.

$$\sum_{k=1}^{\infty} \alpha_k = \infty \quad \text{and} \quad \sum_{k=1}^{\infty} \alpha_k^2 < \infty \quad (3.38)$$

The move size used is  $\alpha_n = 1/n$  and satisfies the above conditions

The descent direction is defined as :

$$d^n = \sum_{rs} Q_{rs} \cdot P^{rsn} \cdot \Delta^{rsT} - q^n \quad (3.39)$$

and each component of the descent vector is given by :

$$\begin{aligned} d_a^n &= \sum_{rs} Q_{rs} \sum_k P_k^{rs} \delta_{a,k}^{rs} - q_a^n \\ &= y_a^n - q_a^n \end{aligned} \quad (3.40)$$

where  $y_a^n$  is an auxiliary flow pattern obtained by a stochastic network loading which is based on the set of travel times  $\{tt^n(q^n)\}$

The MSA algorithm as used in the solution of the SUE problem can then be summarised as follows :

**Step 0: Initialisation.**

Perform a stochastic network loading based on a set of initial travel times  $\{tt_a^0\}$ . This generates a set of link flows  $\{q_a^1\}$ . Set  $n=1$

**Step 1: Update.**

Set  $tt_a^n = tt_a(q_a^n)$ ,  $\forall a$

**Step 2: Direction finding.**

Perform a stochastic network loading based on the current set of link travel times  $\{tt_a^n\}$ . This generates a set of auxiliary link flows  $\{y_a^n\}$ .

**Step 3: Move**

Find the new flow pattern by setting

$$q_a^{n+1} = q_a^n + (1/n)(y_a^n - q_a^n) \quad \forall a$$

**Step 4: Convergence criterion.**

If convergence is achieved, stop, otherwise set  $n=n+1$  and go to step 1.

### Extensions of Stochastic Equilibrium

The concept of stochastic equilibrium was extended by several researchers to solve more general problems.

Daganzo (1983) analysed stochastic equilibrium problems over transportation networks with multiple vehicle types. Sheffi and Daganzo (1978b) introduced the concept of *hypernetworks* (which can be also applied in the deterministic UE approach) to analyse all the choice dimensions as a traffic assignment problem. The hypernetwork includes the links and nodes representing streets and intersections as well as dummy nodes and

links representing the various travel choices. Thus the various alternatives concerning mode, route destination etc are considered as paths made up of links characterised by disutilities. Mirchandani and Soroush (1987) allow the transportation network itself to be stochastic. In their generalised model, the travel time on each route is random and each traveller perceives, possibly inaccurately, a travel time probability distribution for each route which may vary from traveller to traveller. More recently Safwat and Magnanti (1988), based on the stochastic equilibrium principle, developed a combined trip generation, trip distribution, modal split and trip assignment model which was implemented by Safwat and Walton (1988) in the urban transportation network of Austin, Texas.

### 3.7 Summary

This chapter has provided the necessary information needed for the analysis of traffic assignment and reviewed some of the methodologies for solving the static traffic assignment problem.

The assignment process deals with the distribution of an origin-destination matrix into alternative routes of an urban transportation network. In the static assignment approach the O-D trip rates are assumed to be constant during the period of analysis, and demand and traffic are assumed to be uniformly distributed within the modelled time interval.

The solution methodologies are developed with regard to a graph representation of the road network, which however is not unique. The choice of the representation depends on the level of detail at which the network is modelled which in turn depends on the availability of data and the analysis budget. The graph representation of the network includes nodes and directed links, the latter being associated with some impedance that affects the flow using it. The level of service related to a link of the network is represented by a link performance function expressing the relationship between travel time and traffic flow.

All route choice models assume that drivers minimise their travel time. However the travel time needed to traverse a link is an increasing function of the demand using that link. Thus, since congestion increases with flow and trips are discouraged by congestion, this interaction can be modelled as a process of reaching an equilibrium between congestion and travel decisions.

Two types of equilibria are discussed :

- i) the deterministic user equilibrium which assumes that travellers know all the link travel times in the network with certainty and make consistently the correct decisions and,
- ii) the stochastic user equilibrium which assumes that each traveller may perceive a different travel time and act accordingly

The two most widely used heuristic equilibration methods for solving the network equilibrium model are discussed. These are iterative techniques which involve repetitions of all-or-nothing network loading. Such a loading is executed by assigning the flow between each O-D pair to the minimum-travel time path connecting this O-D pair. The heuristic techniques discussed are :

- i) the capacity restraint method which is not guaranteed to converge, and
- ii) the incremental assignment method which may converge to a nonequilibrium solution.

The inadequate performance of these heuristics has motivated the development of the equivalent minimisation approach. To demonstrate that the solution of this program is equivalent to the solution of the UE conditions, it is shown that the equilibrium equations are in fact the first order conditions of the program. This guarantees that the equilibrium conditions hold at any stationary point of the program. It is also shown that this program is strictly convex, meaning that it has only one stationary point which is the minimum. Therefore instead of solving the equilibrium equation directly, the equilibrium flows can be determined by minimising the equivalent minimisation program.

The Frank and Wolfe algorithm is applied in order to solve the UE equivalent minimisation program. The resulting algorithm, as applied in traffic assignment, has a form similar to the heuristic equilibration techniques; it is an iterative procedure which involves an all-or-nothing network loading and an one dimensional optimisation at each iteration. The number of iterations needed is mainly a function of the congestion over the network.

The chapter also discusses the topic of stochastic user equilibrium assignment. In the SUE link travel times are assumed to be both random and flow dependent. The randomness of these travel times stems from the variability in their perception by motorists, while the flow dependency is due to the congestion phenomena.

An essential procedure used for the solution of the SUE problem is the stochastic network loading mechanism, which is the assignment procedure in which link travel times are random but not flow dependent. Three SNL algorithms were presented. The

first is based on a discrete-uniform distribution, while the second and third on the multiple choice models presented in the previous chapter. Thus the **STOCH** algorithm is a logit model for stochastic network loading where each path is included in the subset of used paths (reasonable) only if none of the links along this path takes the driver back toward the origin. The algorithm is very computationally efficient but its use is limited in some respects as it tends to assign too much flow to overlapping paths. The third algorithm is based on the probit model formulation and overcomes the limitations of the **STOCH** algorithm but requires higher computational effort and therefore is not practical for large scale applications.

The **SUE** assignment problem is formulated as an unconstrained minimisation program the solution of which is the **SUE** flow pattern. The minimum of this program is unique although the program is not necessarily convex.

The method of successive averages is applied in order to solve the equivalent minimisation program. This method involves repetitions of a stochastic network loading mechanism and uses a predetermined sequence of step sizes so that the objective function need not be evaluated at any stage of the algorithm.

# 4 peak period work trip scheduling

The following text is extremely faint and illegible. It appears to be a list of items or a detailed schedule, possibly containing names, dates, and times, but the characters are too light to be transcribed accurately.



## **Objective**

The purpose of this chapter is to demonstrate the characteristics of peaking in transportation systems, and then to discuss the work trip scheduling and the different strategies to alter this scheduling in order to reduce the effects of peaks.

### **4.1 Introduction**

The separation between residential and employment areas has created an important characteristic of the urban transportation activities. This is the considerable movement of commuters (initially widely distributed throughout the urban area) who have to travel to and from the Central Business District (CBD), where the majority of workplaces is located. Furthermore, the practice of most working hours being similar through the whole spectrum of jobs leads to large peaks in the daily profile of transport demand, since people travel between home and work at approximately the same times during the day. This heavy concentration of trips both in terms of time and space gives rise to traffic congestion with its inherent problems and inconveniences, and has become a feature of most large cities.

In conventional network analysis, peak traffic flow is estimated by assigning a certain fraction of the daily demand over the peak period. The assignment procedure is based on the assumption that each traveller tries to minimise his disutility of travel and therefore selects the minimum-travel-time route connecting his origin to his destination. However another dimension of his choice set is the time of the day at which he can make his trip; this dimension is very important since during the course of the peak period, travel times are not constant. In general, travellers prefer hours when routes are relatively uncongested, in order to reduce their travel time and the

inconveniences associated with congestion. However they are usually restricted with respect to the range of the available arrival times at the destination; work trip times, for example, are strongly influenced by high institutionalised working hours. Thus a traveller may have a choice between on time arrival with a long travel time during the peak period, and a late or an early arrival with a shorter travel time outside the peak period. The difference between the actual and preferred arrival time is the *schedule delay*, and was introduced by Kraft and Wohl (1967) as a travel time component mainly associated with transit trips due to the discontinuities in the availability of transit service. Later Wohl (1970) argued that schedule delay should be an essential consideration in forecasting and evaluating peak period congestion.

Furthermore, during peak periods there is a greater variability in travel time and therefore a more uncertain time of arrival. Paine *et al.* (1976) argue that for both work and non-work trips, arriving at the intended time is considered more important than average time and cost, which are generally thought to be the dominant service attributes that affect demand. Therefore departure time choice is an important element which should be taken into account in the transportation demand analysis.

The flexibility of departure time decisions can be seen during temporary disruptions of transportation facilities. Hendrickson and Plank (1984) mention that 65% of CBD commuters reported earlier departure times during a transit strike in 1976 in Pittsburgh, and also that work trip departure times were 19 mins earlier during the reconstruction of a major roadway in Pittsburgh, indicating that individuals possess the opportunity to change departure times.

The modelling of departure time choice may be viewed as expressing the individual's internal trade-off of the perceived loss due to early or late arrival, versus the increased travel time in the peak (which includes effects such as the unpleasantness of driving in dense traffic). Commuters are most satisfied when they arrive at work close to their official work start time. As the commuter arrives increasingly later than the official work start time, the magnitude of the perceived loss increases, representing employment penalties that may be associated with tardiness (e.g., loss in pay, poor reputation, and negative impact on promotion). It is therefore presumed that the penalties for being a few minutes late will be far less severe than those for arriving 15-30 min late. Perceived loss is assumed to increase with early arrival as well, since the commuter could have used the extra time as leisure time at home, which is likely to be valued more than being at the office. Furthermore perceived penalties for late

arrival at work are expected to be greater than perceived penalties for not maximising leisure time at home.

Vickrey (1969) was one of the first researchers who studied travellers' shifting of departure times in order to avoid peak period congestion. He considers 4 different values of time for the commuters; the value of time : i) spent at home ( $w_h$ ), ii) spent at office prior to the desired starting time ( $w_p$ ), iii) spent in the queue ( $w_q = 0$ ) and iv) for time after the desired starting time (i.e. wage rate,  $w_j$ ), such that  $w_j > w_h > w_p > w_q$ . A commuter is then assumed to select that departure time that will maximise the overall value of his time. Vickrey's approach was extended by several researchers to model work scheduling decisions within the utility maximisation context, by estimating econometric logit models of the choice of departure or arrival time with different specifications of the utility functions.

Management policies for urban transportation systems often involve direct or indirect influences on users' departure time decisions. Capacity expansions to roadway facilities may result in additional bunching or peaking of departure times, so that the resulting reduction in congestion is less than would otherwise occur. Staggered and flexible work hours, and peak period transit fares and tolls attempt to directly influence departure time decisions. A better understanding of the factors which influence departure time decisions is therefore of great importance

Thus the next section provides a description of the peak period phenomena with emphasis on the time-of-day variability of peak period flows and travel times in urban transportation networks. In section 4.3 some models of departure time choice are reviewed in order to define the factors influencing work trip scheduling. In section 4.4 alternative strategies for urban transportation management, which involve influences on travellers trip scheduling, are discussed and finally section 4.5 summarises the chapter.

#### **4.2 The peaking nature of traffic flows and travel times**

The main reasons for the transportation peaking problem in urban commuter routes are:

- the high concentration of employment within a relatively small district, known as the Central Business District (CBD), and
- the scheduling of working hours which is such that the majority of employees in a city have to arrive at work within a predetermined short time interval.

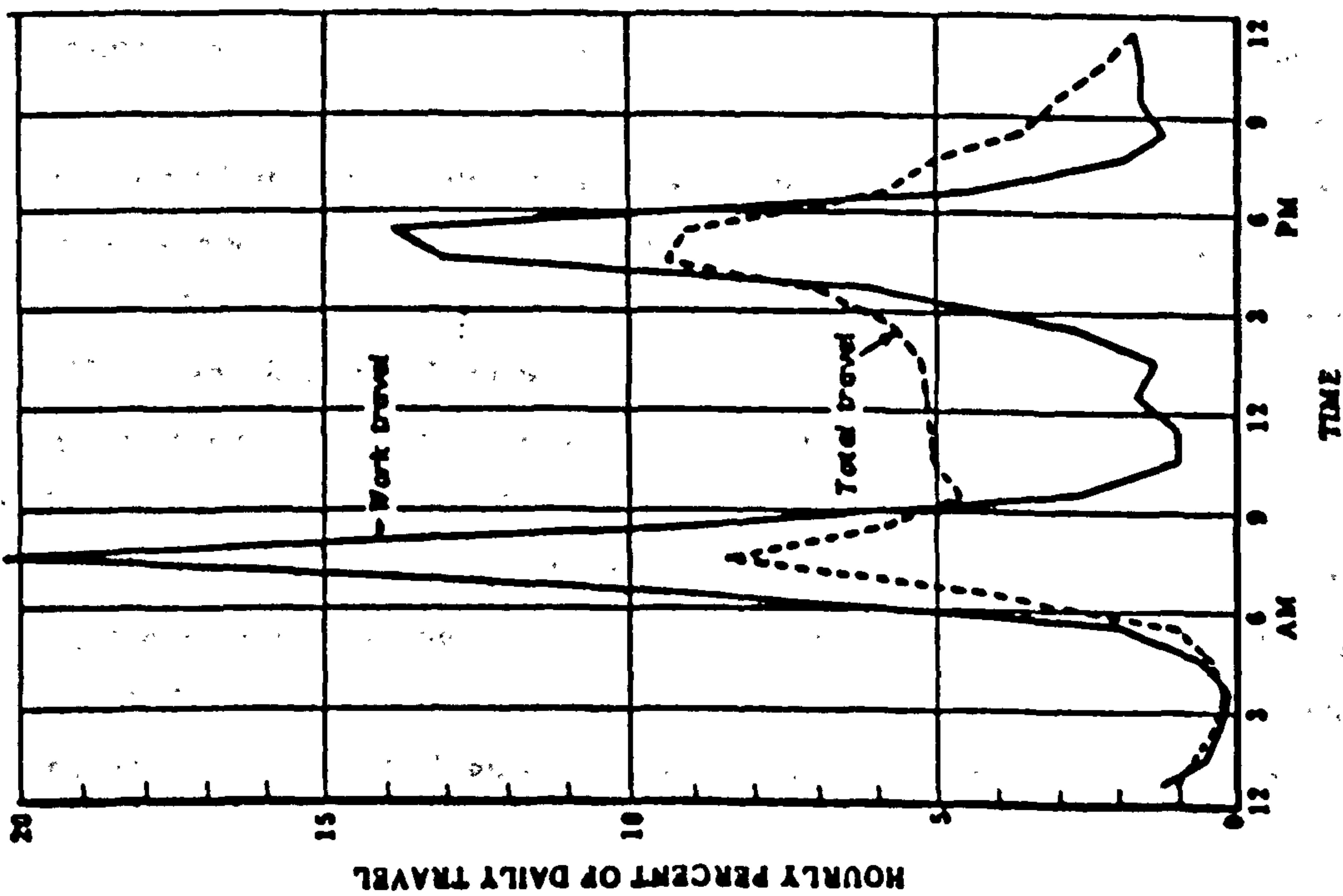
This section presents data on the observed time of day variability of traffic flows and travel times, in order to illustrate and quantify the nature of peaking in highway systems.

#### **Traffic flow variability**

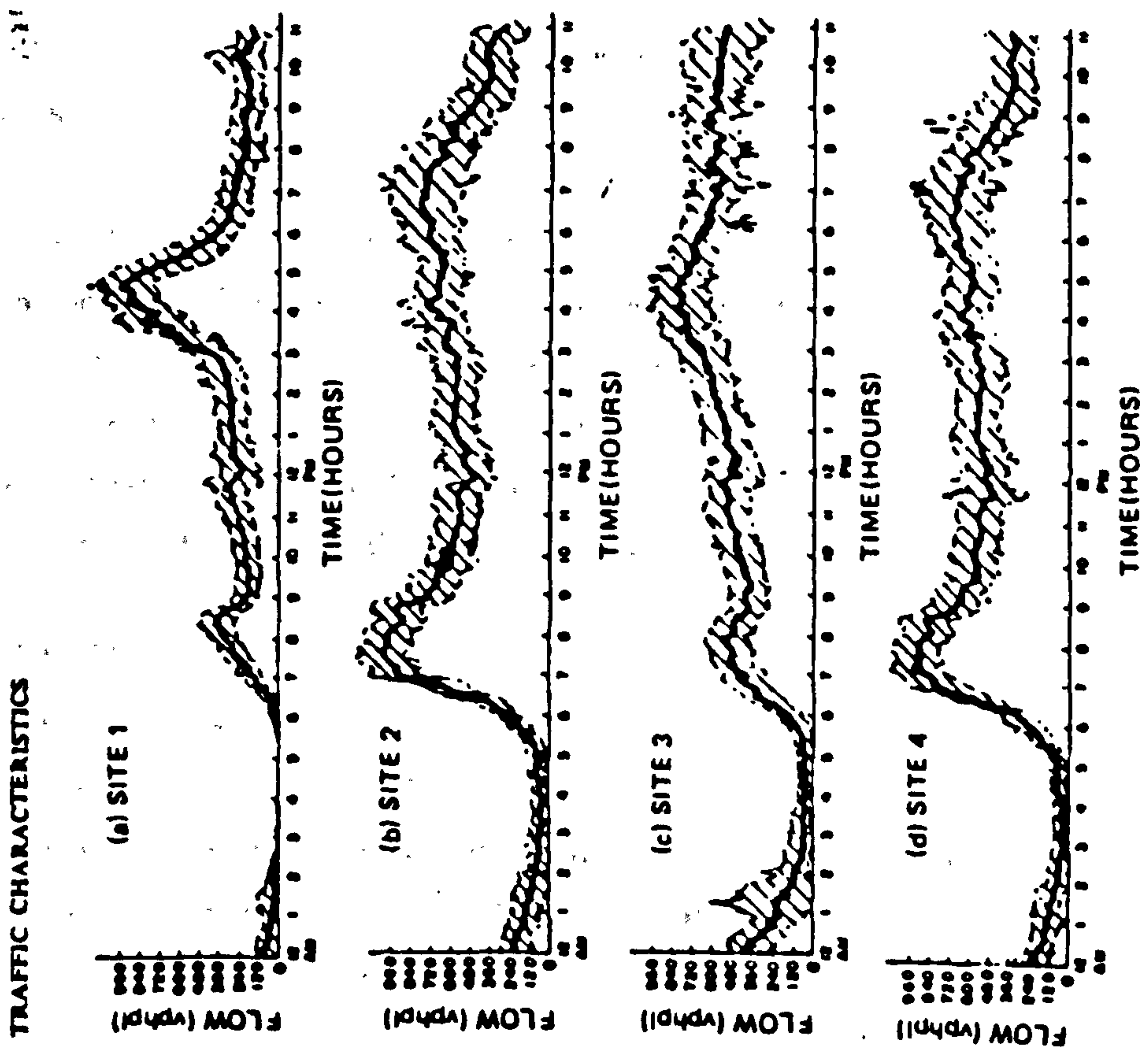
As described in the Highway Capacity Manual (TRB, 1985) traffic volumes vary in both space and time; traffic demand varies by month of year, by day of the week, by hour of the day, and by subhourly intervals within the hour. This section will concentrate on the hourly and subhourly variation in traffic demand, since the accurate estimation of these variations is very important if highways are to effectively serve peak demands without breakdown. A breakdown occurs when the ratio of actual arrival flow to actual capacity of a road section exceeds 1.00, and as a result queues form and consequently delays occur. These effects may extend far beyond the time during which demand exceeds capacity, and cause the time of day variability of travel time.

Figure 4.1 (TRB, 1985) depicts the hourly distribution of vehicle work travel and total travel averaged for the following urban areas: Boston, St. Louis, Seattle, Louisville, Oklahoma City, Colorado Springs, Stockton, and Fall River, Massachusetts. The figure shows the difference between work travel and total travel distributions. Work travel is much more heavily concentrated in the morning and afternoon peaks than the total vehicle travel is. Between 7:00 and 8:00 a.m., 20.2% of the daily work travel occurs, while during the same interval only 8.4% of the total travel occurs. Similarly in the afternoon peak, both the 4:00 pm to 5:00 pm and the 5:00 to 6:00 interval accounts for about 13% of the daily work travel and for about 9% of the total travel.

Traffic analyses mainly focus on the peak hour of traffic volume, because it represents the most critical period for operations and has the highest capacity requirements. The peak hour volume, however, is not a constant value from day-to-day or from season-to-season. Urban routes show very little variation in peak-hour traffic, during which the majority of users are daily commuters. Furthermore, many urban routes are filled to capacity during each peak hour, and variation is therefore severely constrained. This agrees with McShane's and Crowley's (1976) observations who have shown the typical variations by hour of the day, and the repeatability of traffic patterns using data obtained over a 77-day period in metropolitan Toronto. Their observations are illustrated in fig 4.2 where the shaded area indicates the range within which 95% of the observations are expected to fall.



4.1 Hourly distribution of vehicular travel in selected urban areas. (Source: Highway capacity manual, TRB, 1985).



NOTE: (1) SITES 2 and 4 ARE ONE BLOCK APART ON SAME STREET, IN SAME DIRECTION  
(2) ALL SITES ARE TWO MOVING LANES

4.2 Repeatability of hourly traffic variations for four 2-lane arterials in Toronto, Ontario, Canada (Source: Mc Shane and Crowley, 1976)

Figure 4.3 illustrates the substantial short term fluctuation in flow rate that can occur during an hour. It can be seen that the maximum 5-min rate of flow is 2232 vhs/h, the maximum rate of flow for a 15-min period is 1980 vhs/h, and the full-hour volume is 1622 vhs/h. Thus a design for an hour volume would result in congestion for a substantial portion of the hour, which however could not be estimated if the variation of flow within the peak hour is not taken into account. Consideration of the peak rates of flow occurring within the peak hour is important, because congestion due to inadequate capacity occurring for only a few minutes could take substantial time to dissipate because of the dynamics of breakdown flow.

#### **Travel time variability**

The travel time characteristics of commuting trips have been analysed in several studies, such as the ones by Herman and Lam (1974) and Richardson and Taylor (1978). However, few studies report on changes in travel time characteristics during the course of the peak period. In the following paragraphs these studies will be used, in order to illustrate the time of day variability of travel time.

Smeed and Jeffcoate (1971) used a record of travel times for journeys between Windsor Road, Bray, and Flaxman Terrace, London WC1, to support their argument that journeys by road take much longer at some time of the day than at others. Figure 4.4a shows the time taken for each of the 96 journeys that recorded the time of starting from Bray and the time of arrival at Flaxman Terrace during the period June 1969 to May 1970. This journey time is plotted against the time of starting from Bray. It is clear from the figure that the time taken for the journey depends on starting time to an appreciable extent. The discrepancies between journey times for journeys started at about the same time are explained: accidents cause long delays and during holidays lower travel times are experienced. Between 7.00 am and about 7.50 am the average journey time varied approximately linearly with the starting time. During this period the journey took about 0.63 minutes longer for every minute after 7.00 am that the journey started. There was little trend in average journey time for journeys starting between 7.50 am and 8.30 am. Between 8.30 am and 9.30 am there was a reduction in average journey time of about 0.24 minutes for every minute the journey started after 8.30 am. For the morning journey into London the shortest time was 44 min and the longest 116 minutes, so that the longest time in 96 journeys was 2.6 times the shortest. On average, the time of starting that resulted in longest journey times, i.e. between 7.50 am and 8.30 am, resulted in journey times about 60% longer than those which started between 7.00 and 7.15. Figure 4.4b gives the corresponding journey times for

TRAFFIC CHARACTERISTICS

LOCATION MIDDLE LANE OF SOUTHBOUND INTERSTATE 39W IN MINNEAPOLIS, MINNESOTA, 4 MILES SOUTH OF CENTRAL BUSINESS DISTRICT, THREE LANES (ONE WAY)

TIME WEEKDAY MORNING, METERED, AUGUST, 1983

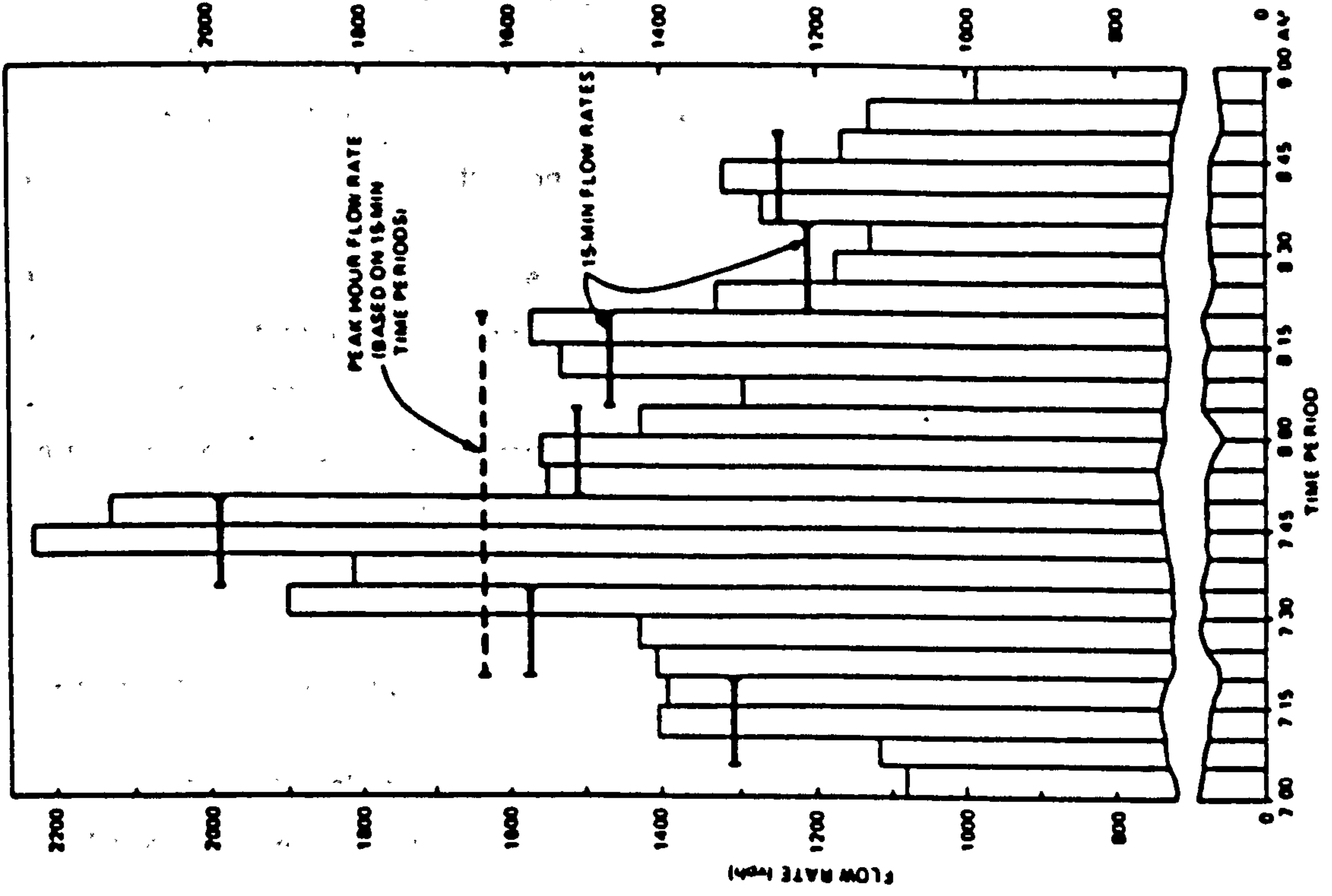


fig. 4.3 Relationship between short-term and hourly flows (Source: Highway capacity manual, 1985).

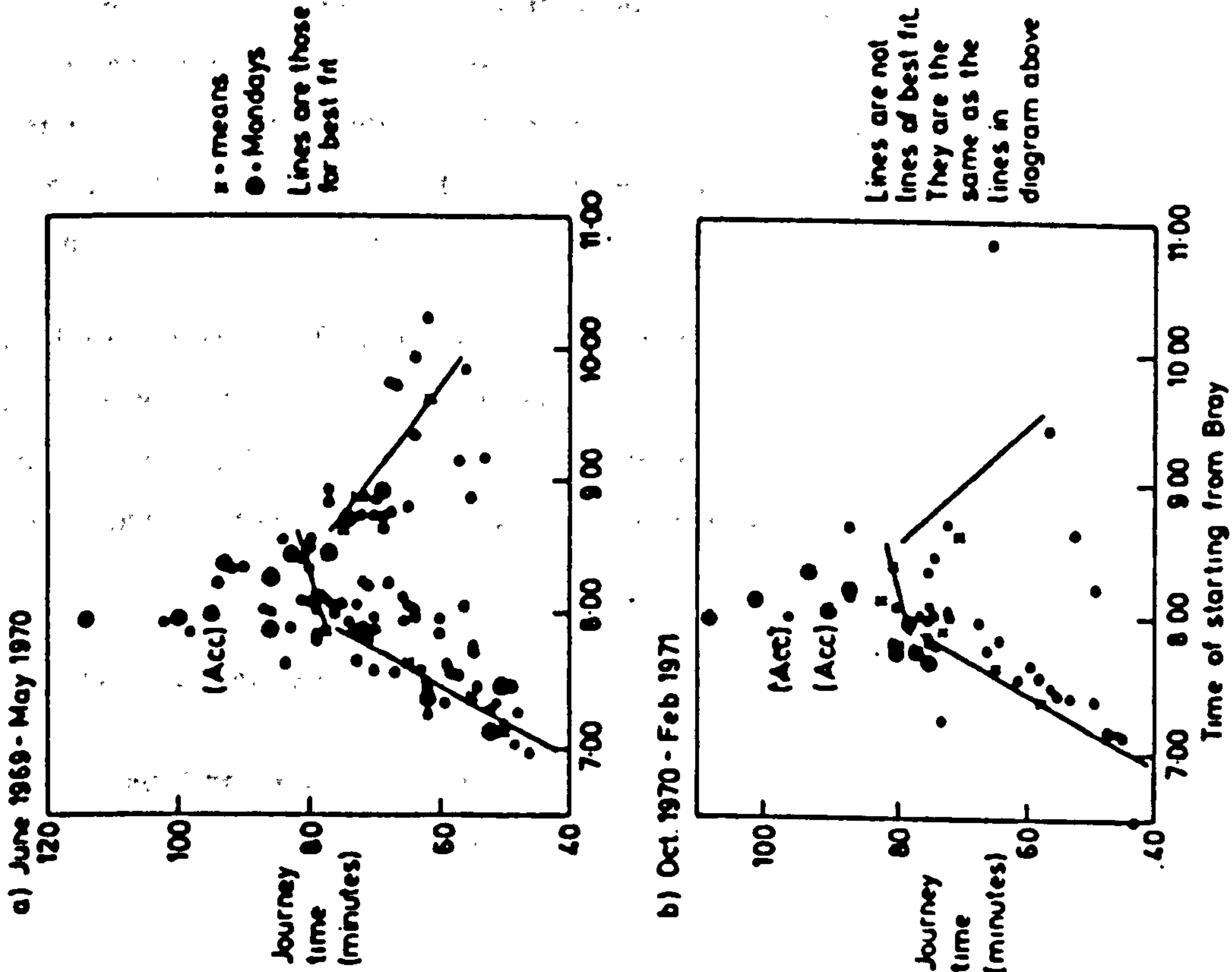


fig. 4.4 Journey time, Bray to London WC1. (Source: Smeed and Jeffcoate)

another period, from October 1970 to February 1971. The lines shown in this figure have not been calculated as the best fit lines for the data; they are the lines shown in fig. 4.4a, and it is seen that there is little difference between the trends given by the two sets of data, at least for the early starting times.

Hendrickson *et al.* (1981) have recorded the travel times from five zones of Pittsburgh to the CBD. They found that the difference between the peak period travel time and the shortest travel time ranged from 3 to 16 minutes, representing between 25% and 52% of the minimum travel time. Their data is illustrated in fig. 4.5. For Glenshaw and Shadyside, there is no sharp peak travel time, but rather a plateau representing a peak period travel time. The routes from these two areas to CBD seem to have sufficient capacity since no major bottlenecks exist. However, Bethel Park, Greentree and Monrieville exhibit definite peaking patterns.

For the same data set, Hendrickson and Plank (1984) have further observed that the pattern of travel time peaking was quite regular in each area, with the maximum travel time occurring at about the same time on each day of observation. They suggested that the increase in congestion to the maximum travel time and the subsequent decline can be represented by a quadratic function in which extra or congestion travel time was a function of departure time from the suburb. (fig. 4.6). They also found that, despite the scatter of measurements shown in fig. 4.6, the variability in travel times for a given departure time was relatively low since the average ratio of the standard deviation to average travel time (i.e. the coefficient of variation) was 0.13 over all routes and departure times. This suggests that, with experience, commuters can fairly accurately predict the amount of time required to travel to work on any given day with a particular departure time, implying that travel time peaking occurs in a fairly regular pattern from one day to the next. This agrees with McShane's and Crowley's (1976) observations as shown in figure 4.2.

The variability of traffic flow patterns and travel times during the course of the peak period, is associated with the time varying levels of the demand for travel, and therefore with the individuals' decisions on departure time selection. A review of some studies on departure time choice is given in the next section.

### 4.3 Work Trip Scheduling

Work trip scheduling and therefore departure time decision is directly related to the desired arrival time at work and thus to the official work starting time. Arriving early



**Text cut off in original**

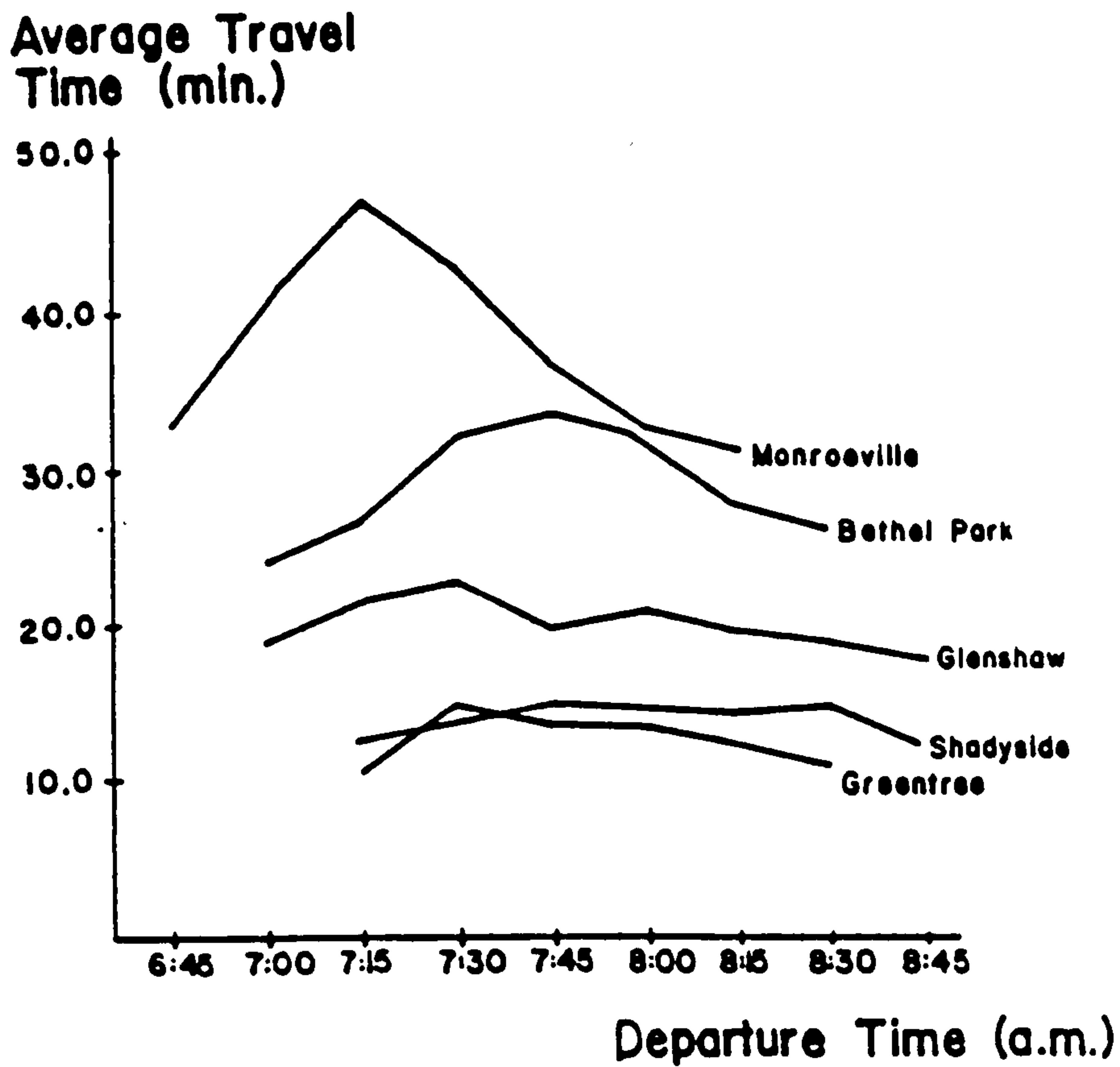


fig. 4.5 Average travel times by departure time for five zones in Pittsburgh, PA.  
(Source: Hendrickson et al, 1981)

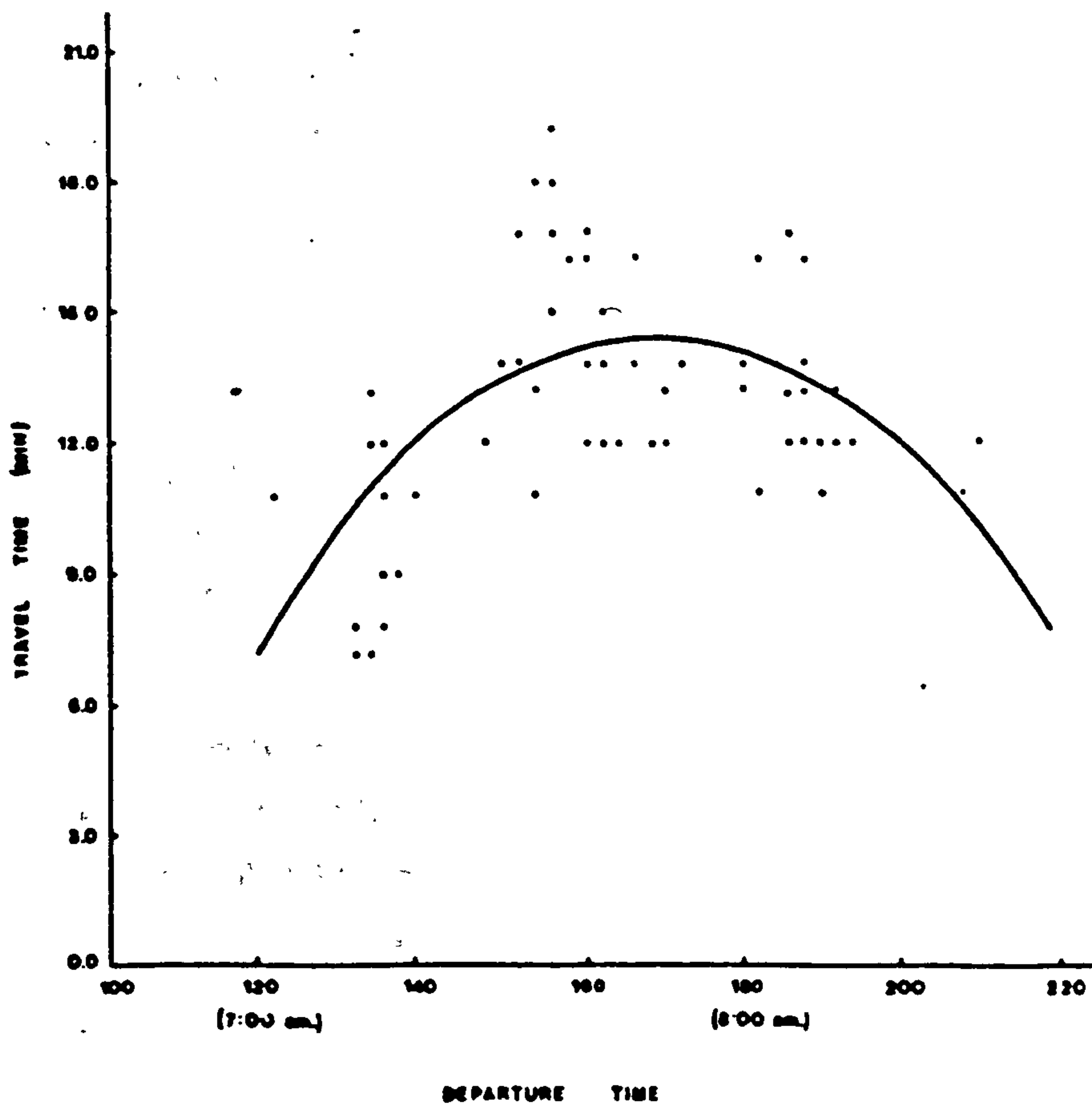


fig. 4.6 Travel times at different departure times from Greentree to the Pittsburgh CBD  
(Source: Hendrickson et al, 1981)

at work is likely to involve some time wasted, or at least less productively used, and therefore decreases utility; arriving late has, for most workers, more severe consequences. Furthermore travel time, another factor that decreases utility, is not constant during commuting periods, but depends on departure time. This time of day variation of travel time is in turn influenced by the level of demand using the transport facilities, and thus depends on the departure time choice of the individuals. It is, therefore, the trade-off between scheduling considerations, as represented by the early or late schedule delay, and travel time which is crucial for studying the impact of scheduling behaviour on congestion.

Recently, the realisation of the importance of considering the time dependent nature of travel demand and transportation level of service has led to the increased empirical research into the area of departure time decision. Several researchers have studied the departure time decision separately or in relation with the mode choice decision. In this section some of these studies are discussed with the aim to introduce the factors that influence work trip scheduling.

Cosslett (1977) estimated a multinomial logit model for the departure time decision. Using data for the Urban Travel Demand Forecasting Project (UTDFP) sample; he examined the individual's trade-off between mean travel time, schedule delay and probability of arriving late in the case of trips to work by car. In his work schedule delay is defined as the difference between the actual mean arrival time and the official work start time. This differs from the classical definition of schedule delay as the difference between the desired mean arrival time and the actual mean arrival time due to discontinuities in the availability of transit service. Cosslett used the following utility form :

$$U(t) = -\alpha.y(t) - \beta.E - \gamma.L(t) \quad (4.1)$$

where :

$t$  = the individual's arrival time at work

$y(t)$  = the on-vehicle time corresponding to arrival time  $t$

$$E = \begin{cases} T_1 - t & \text{if } T_1 \geq t \\ 0 & \text{if } T_1 < t \end{cases}$$

i.e. early arrival time at work, in minutes

$T_1$  = the individual's official work start time

$L(t)$  = the probability that the individual will arrive late for work, if he plans to arrive at time  $t$

To calculate  $L(t)$ , on vehicle time was replaced by a normal random variable with mean  $y(t)$  and standard deviation  $\sigma = a[y(t) - y(0)]$ , where  $y(0)$  is the off-peak travel time and  $a$  is a constant, assuming that the standard deviation should be approximately proportional to excess travel time due to congestion. ( $a$  has taken several values, however the results reported by Cosslett were for  $a = 0.2$ ). The arrival time alternatives were discrete intervals at the minutes ranging from 40 minutes early to 15 minutes late, and all the estimated coefficients were significant, particularly the late and early measures. The relative disutility of congestion time to schedule delay,  $\rho = \alpha/\beta$ , is 1.6 for the travellers who travel alone, implying that these commuters are willing to incur 0.62 minutes of travel time to avoid arriving one more minute early. It was also found that these travellers are willing to travel for 2.14 minutes more in order to avoid a 10% increase in the probability of arriving late. Another conclusion from this study was that the response of shared-ride travellers to congestion time is much smaller and is consistent with zero reflecting the greater difficulty in scheduling shared rides.

Following Cosslett's study, Small (1982) also analysed the scheduling of work trips, using discrete time periods and assuming mode choice as fixed, and estimated a departure time model using the same UTDFP data set. His formulation of the model is:

$$W(s) = \beta_1 RPTR15(s) + \beta_2 RPTR10(s) + \beta_3 TIM(s) + \beta_4 SDE(s) + \beta_5 SDL(s) + \beta_6 DIL(s) \quad (4.2)$$

where :

- $W(s)$  = utility for time period  $s$
- $RPTR15(s)$  = round-off bias variables for reporting of intervals of 10 and 15 minutes of schedule delay (1 if round-off present; 0 otherwise)
- $RPTR10(s)$  = round-off bias variables for reporting of intervals of 10 and 15 minutes of schedule delay (1 if round-off present; 0 otherwise)
- $TIM(s)$  = mean travel time
- $SDE(s)$  = schedule delay early (schedule delay if early arrival; 0 otherwise)
- $SDL(s)$  = schedule delay late (schedule delay if late arrival; 0 otherwise)
- $DIL(s)$  = late dummy variable (1 if schedule delay late; 0 otherwise)

Small estimated that the *marginal rate of substitution* between travel time and early schedule delay was 0.61, which is consistent with Cosslett's findings. He also concluded that late arrival is more onerous, with arrival beyond a *margin of safety* carrying a penalty equivalent to 5.5 minutes of travel time, plus 2.4 minutes for every minute late, implying that travellers are more sensitive to late arrival than early arrival by a factor 4. He extended his basic model by including other variables such as family status,

transportation mode, occupation and work-hour flexibility. Figure 4.7a and 4.7b illustrate the relationship between the perceived penalty (expressed in minutes of equivalent travel time) and the arrival time for two alternative model formulations. His conclusions are summarised below:

- Carpoolers are less sensitive to travel time considerations than car travellers.
- White collar workers are less sensitive to late arrival.
- Late arrival is less onerous for workers who report some flexibility, (depending on the level of flexibility, as shown in figures 4.7a and 4.7b).

Abkowitz (1980,1981a,1981b), in his study of service reliability and work-travel behaviour, considered the interdependence of the mode and departure time decisions, also taking into account the travel-time uncertainty in these travel decisions, and expanded the choice set to include a study of automobile, transit, and carpool commuters. He estimated a logit model and incorporated variables such as work arrival time flexibility, occupational characteristics, income, actual mode chosen, age, sex, location of home and work, travel time, and expected loss due to early or late arrival. The expected loss was estimated using the distribution of individual arrival times at work, and the arrival time loss function  $l(t)$  to express how travellers perceive varying degrees of loss associated with different arrival times at work. The arrival-time loss function,  $l(t)$ , (loss associated with arrival at time  $t$ , expressed in units of utility), is illustrated in figure 4.8. The slope of the lateness function is roughly 3 times greater than the slope of the earliness function. This finding agrees with the expectation that late arrival is much more onerous to travellers than early arrival at work, and is quite close to Small's estimates that the marginal rate of substitution between early and late schedule delay is about 4.

It is important to note that the earliness and lateness functions intersect the x-axis at different points resulting in some question as to which function is appropriate to use for the overlapping range. The reason for the overlapping functions is due to the fact that different individuals have different preferred arrival times, and that when this information is aggregated, the aggregate estimations produce overlapping functions

The conclusions of his study are summarised below :

- Departure time and modal choice appear to be interrelated in a way that suggests structuring departure time and modal choice decisions as a nested choice rather than a joint choice.
- It appears that arrival time consideration affect modal choice decisions as well as departure time decisions.

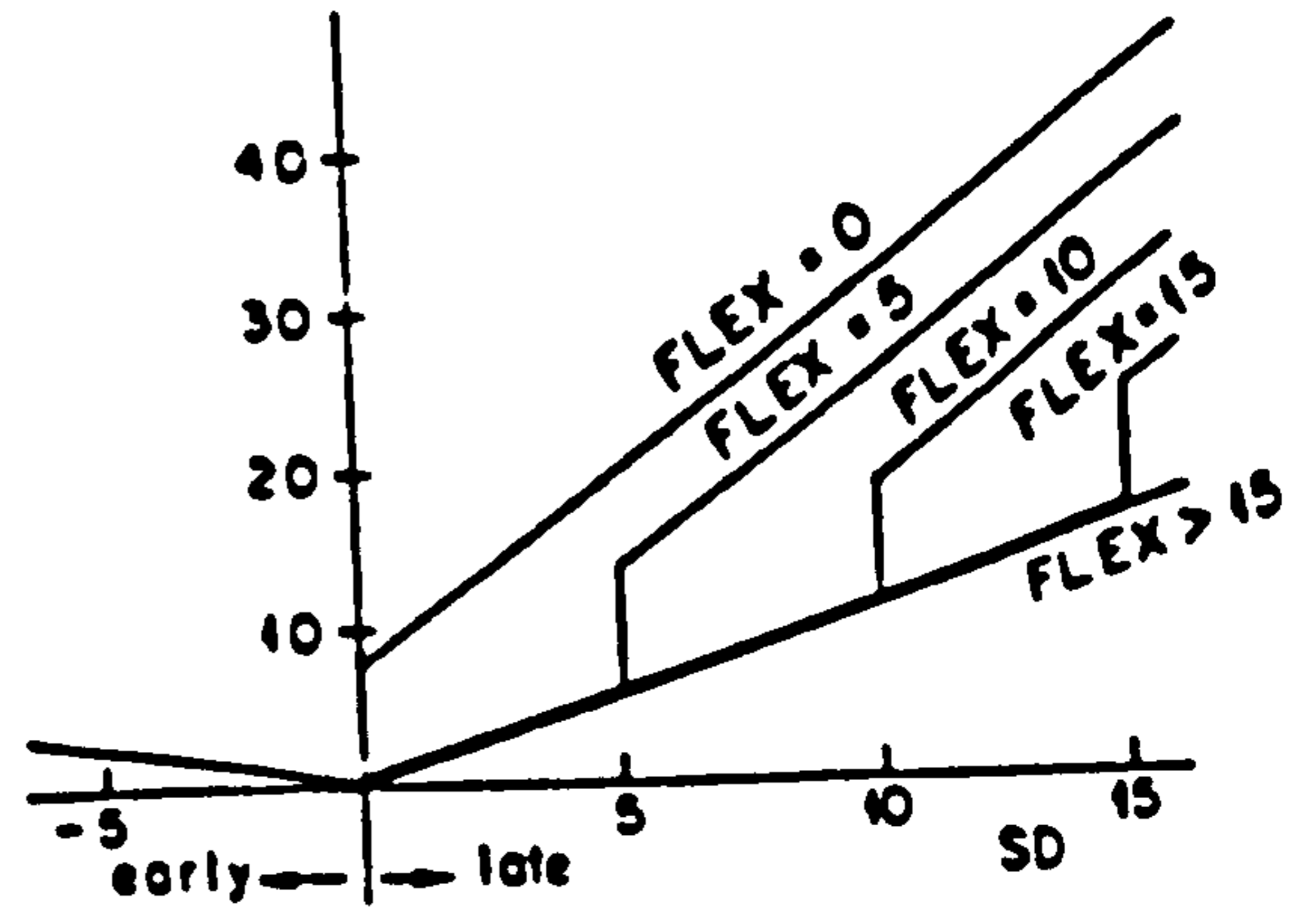
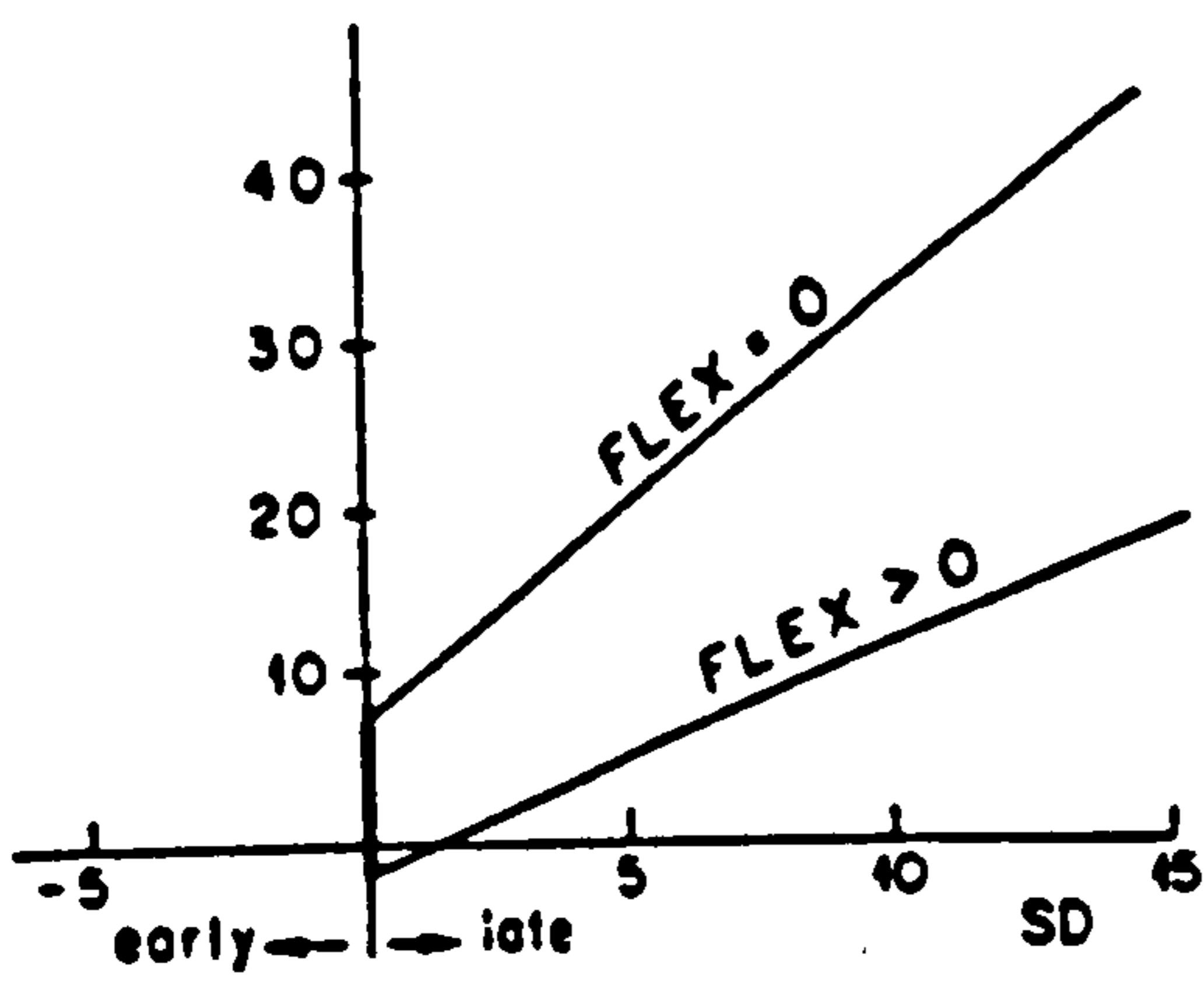


fig. 4.7a Disutility of schedule delay: model 3. fig. 4.7b Disutility of schedule delay: model

(Source: Small, 1982)

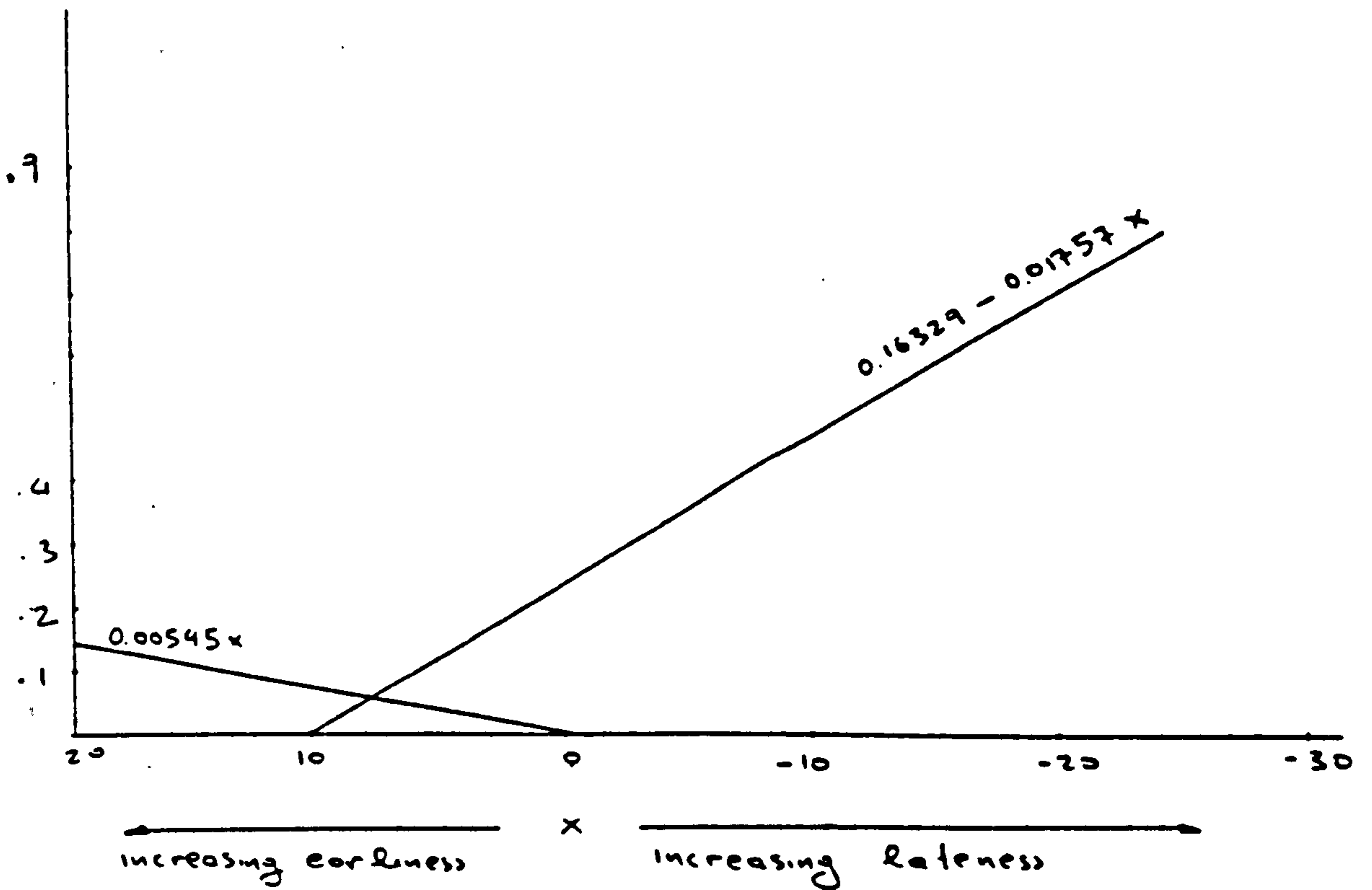


fig. 4.8 Complete Arrival Time Loss Function  
(Source: Abkowitz, 1980)

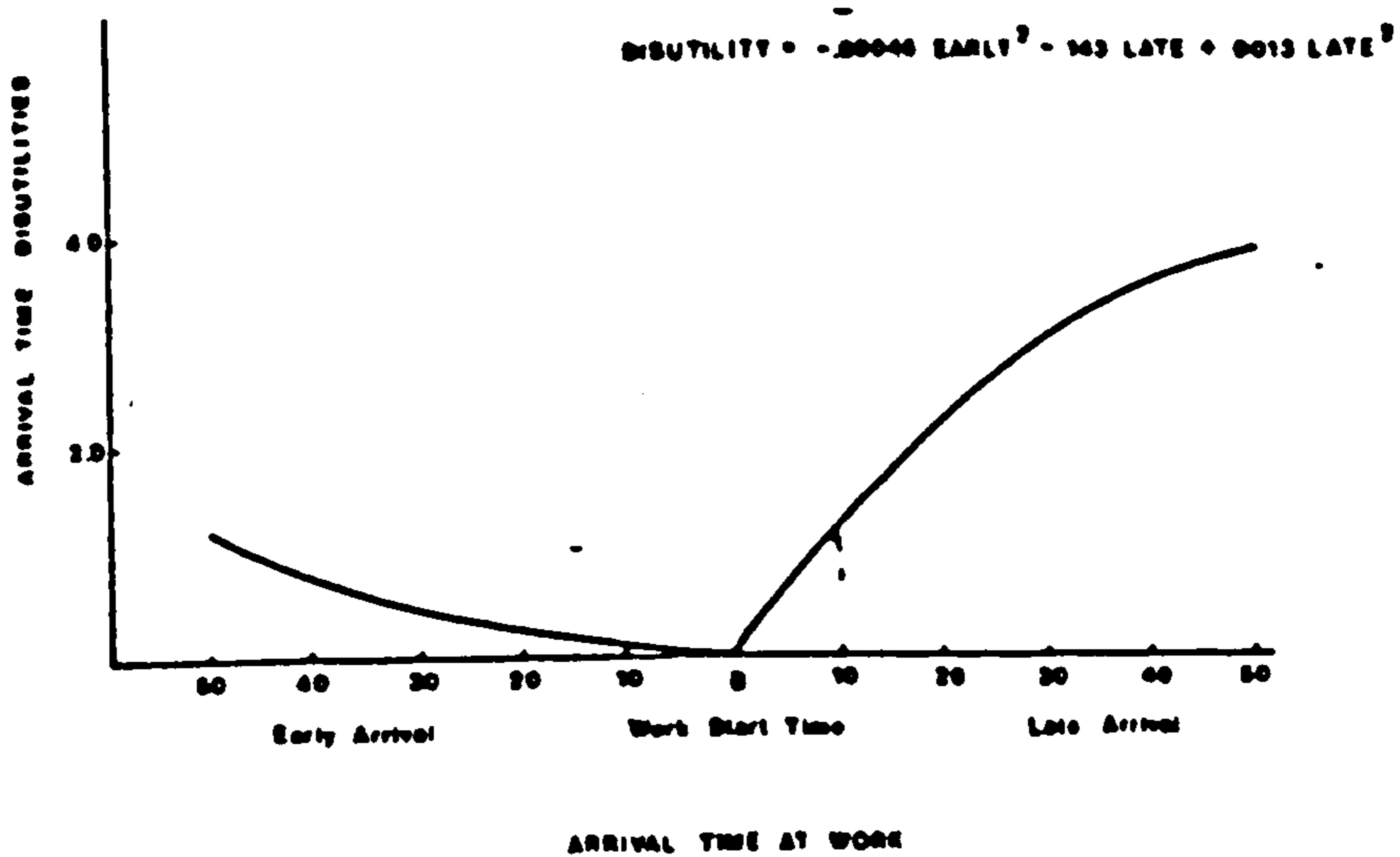


fig. 4.9 Relative disutility of different arrival times at work.  
(Source: Hendrickson and Plank, 1984)

- The availability of a flexible work schedule is important for people planning to arrive exactly on time, and extremely important for those planning a late work arrival.
- Auto travellers are more likely to plan on arriving at work exactly on time, while bus travellers are not likely to depart so as to arrive extremely early for work.
- Individuals employed in a professional, technical, management or administration job typically avoid departure times that result in early arrivals at work.
- Low income workers tend to arrive at or slightly before the official work start time.
- Older workers tend to depart so as to arrive earlier than the official work start time.

Similar conclusions are drawn from the work of Jovanis and Moore (1984) and Moore *et al.* (1984) who have developed a conceptual structure to understand the work arrival time process of workers with flextime, which was tested with actual data. They also found that there are earlier arrivals among workers living at greater distances from the workplace, and that family structure and the associated socioeconomic variables are also very important in explaining work arrival time choice behaviour. Thus the greater levels of family related constraints, as shown by the presence of a working spouse and/or young children, indicate workers that have less flexibility to alter their arrival times with flextime.

Hendrickson and Plank (1984) used a data set on commuting travel, gathered in Pittsburgh, in order to analyse, mode and departure time choices. They treated these choices as a simultaneous interactive decision based upon maximisation of individual traveller's utility or satisfaction with each alternative mode and departure time combination. The utility function they propose includes free flow travel time, delay time, access and waiting time (for transit users), travel cost, and early and late schedule delay. The implied values of time based upon the estimated model coefficient values are summarised in table 4.1 for the average income in the sample. The dollar amounts presented by this table appear reasonable both in terms of their magnitude and relationships. The high cost associated with access to transit may be explained by the fact that in four of the five study areas, access was mainly along unpaved roadway shoulders with no sidewalks. The high late time cost represents the significant penalty which individuals perceive to be associated with late arrival at work. An implicit value of \$2.52 was placed upon five minutes late, while a value of \$4.79 was placed upon being ten minutes late. This suggests that individuals would be willing to pay about

\$2.52 to avoid five minutes late, on average, and \$4.79 to avoid being ten minutes late. Generally, late time is regarded much more onerously than is early time. The complete arrival time cost relationship, including both early and late arrival time, projected by these final model coefficients, is graphed in figure 4.9. This graph indicates the relative disutility of different arrival times at work. The shape of the late time disutility is surprising, since it was expected that disutility would increase more rapidly with later and later arrival after the official work start. What was found, however, contradicts this. Based on the estimated model coefficients, the disutility associated with these later arrivals increases only slightly, implying that once an individual is quite late (approximately 1hr) in arrival, additional late time has little effect.

**Table 4.1: Values of time based upon estimated model coefficients**

Time Variable	Average value of time
Free flow travel time	\$1.71/hr
Congestion time	\$4.50/hr
Access time	\$20.35/hr
Wait time	\$17.14/hr
Late time	\$2.52 for being 5 min late \$4.79 for being 10 min late
Early time	\$0.04 for being 5 min early \$0.15 for being 10 min early

Hendrickson and Plank used their model to forecast the demand effects of a variety of policies on both modal split and travel peaking patterns. They found that departure time decision seems to be more elastic than the choice of mode, and further concluded that this flexibility in departure time choice has some important implications for transportation planning, as :

- Congestion relief due to capacity increases or transit service improvements is likely to be overestimated since, reduction in congestion will cause shifts into travel during the most congested period.
- Departure time flexibility offers a *latent* capacity in transportation facilities which may be useful during short term disruptions such as transit strikes or reconstruction of heavily used facilities.



- Time dependent tolls are likely to be useful tools in transportation system management, particularly if the tolls are more greatly differentiated than a simple peak/off-peak differential.

Having discussed the concept of the departure time decision and the trade-offs involved in work trip scheduling, the next section will focus on the analysis of some strategies which can be used to influence this scheduling, as a means of reducing congestion. These strategies do not involve expansion of existing capacities of transportation facilities and therefore are not associated with high investment costs.

#### **4.4 Managing the Work Trip Scheduling**

Due to the large peaks in the daily profile of transport demand, the transport infrastructure tends to become saturated during the morning and afternoon peak hours, while at other times during the day, roadways and public transport operate with loads well below capacity.

During the last two decades there has been an interest in methods to manage peak period demand rather than increase the capacity of transportation facilities. This interest has arisen because of the increasing construction cost of new facilities and the difficulties in finding acceptable locations for them.

Sellinger (1977) suggests that

- staggered work hours,
- flexible work hours,
- reduced transit fares for off-peak transit users, and
- increased peak hour commutes tolls on access routes to the city,

are techniques which can be used to spread the work travel demand over a longer period of the day, and consequently to reduce the size of the peak demand and make a better use of the existing facilities.

*Staggered hours* programs involve shifting fixed standard, work schedules to earlier or later time periods without changing the length of the workday. Employees have still to be at work by a specified time and leave at a specified time. Under this work schedule, different groups of workers within an organisation, or between organisations within an employment area, have to arrive at work at staggered times (e.g. 7:00 a.m., 7:15 a.m., 7:30 a.m. etc.), therefore resulting in a spreading of the total work demand within a longer period. The reduction of the congestion therefore depends on the degree to

which the new schedules are spread and there is no element of traveller choice (with respect to work starting time) involved in such a program.

*Flexible hour* programs, on the other hand, allow each employee some freedom in determining work schedules, since employees are permitted to set their own daily starting and quitting times within pre-established limits, as long they work the total weekly hours. If travellers make sufficient use of flexible schedules, work trip demand may be spread over a long enough time period (depending on the degree of flexibility) to reduce peak demand significantly with the immediate benefit of increasing efficiency in transport facilities.

Kemp (1977) argues that the best form of work schedule adjustment, as far as the beneficial transportation consequences is concerned, appears to be the flexible working hours. However they are more difficult to promote to employers, and most of the companies adopting flexible time do so only after an initial period of work hour staggering.

The positive impacts of variable work schedule on transportation can be seen from the outcomes of the implementation of such strategies. An extensive description of these outcomes in different cities in North America is given in the report 'Alternative Work Schedules: Impacts on Transportation' (TRB, 1980). In the following paragraphs the actual changes in the profile of peak demand and peak travel times that have occurred as a result of the implementation of staggered and flexible work hours, in some case studies taken from the above mentioned report, are presented.

An evaluation of the effects of a variable work hours program was conducted in the Queen's Park area of Toronto, involving 11000 government employees in October 1973. Sixty eight percent of the employees were assigned to staggered hours and another 23% were placed on flexible work hours. As shown in figure 4.10, before the implementation of the program 90% of the workers arrived at work during the 8:00 to 9:00 peak hour, whereas after the implementation 52% arrived at work before 8:00 and 43% arrived between 8:00 and 9:00.

Figure 4.11 depicts the time distribution of traffic volumes crossing a screenline corresponding to the boundaries of the Ottawa central area across which a high percentage of peak-period traffic is bound to or from CBD jobs. The morning peak 15-min volume was 6% lower after flexible hour schedules were introduced and occurred 15 mins earlier. It should be noted that the total morning peak-period volume between 7:00 and 9:30 increased by 10% in the time between the before and after

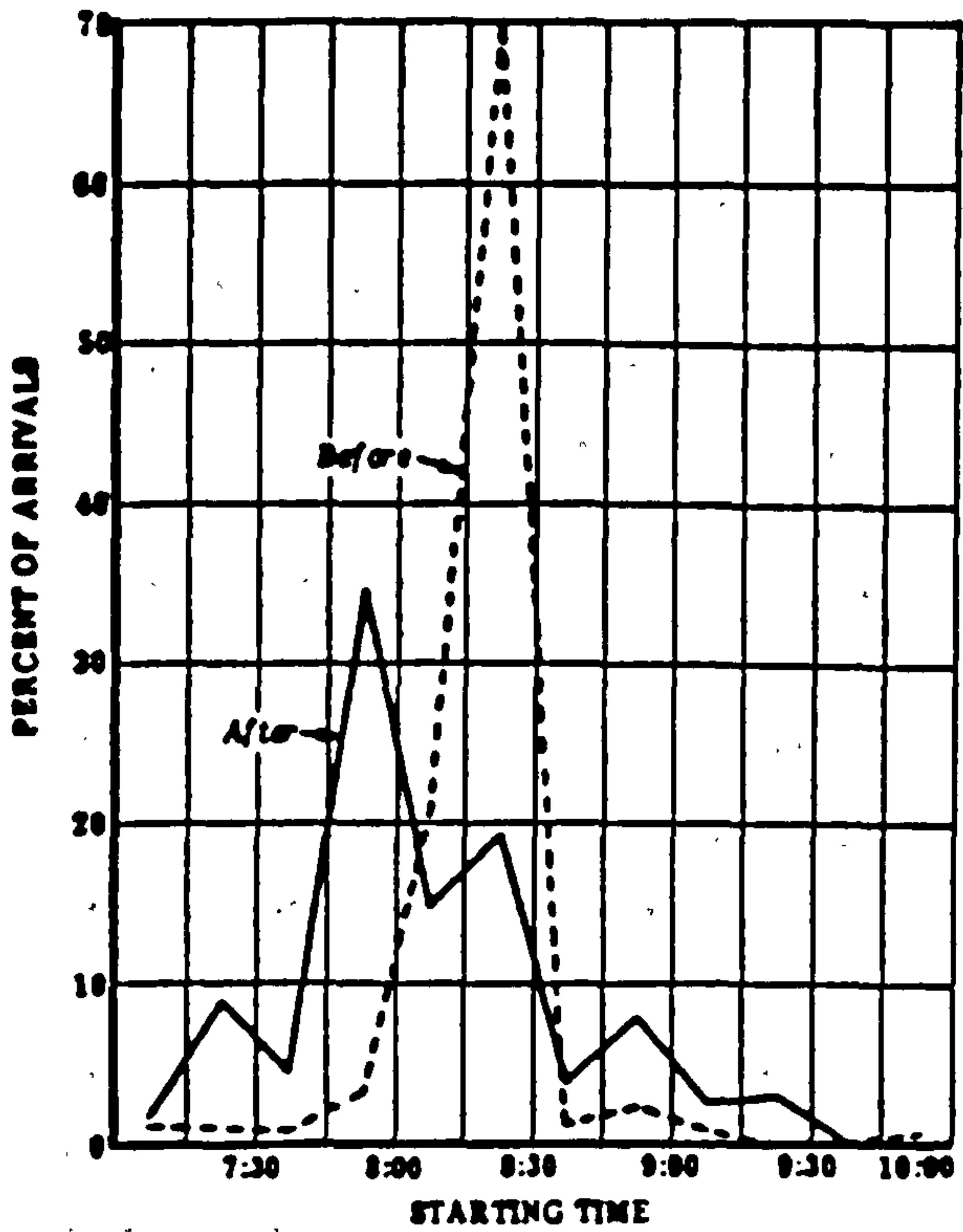


fig. 4.10 Impact of flexible work hours on arrival time of government employees in Queens Park, Toronto.

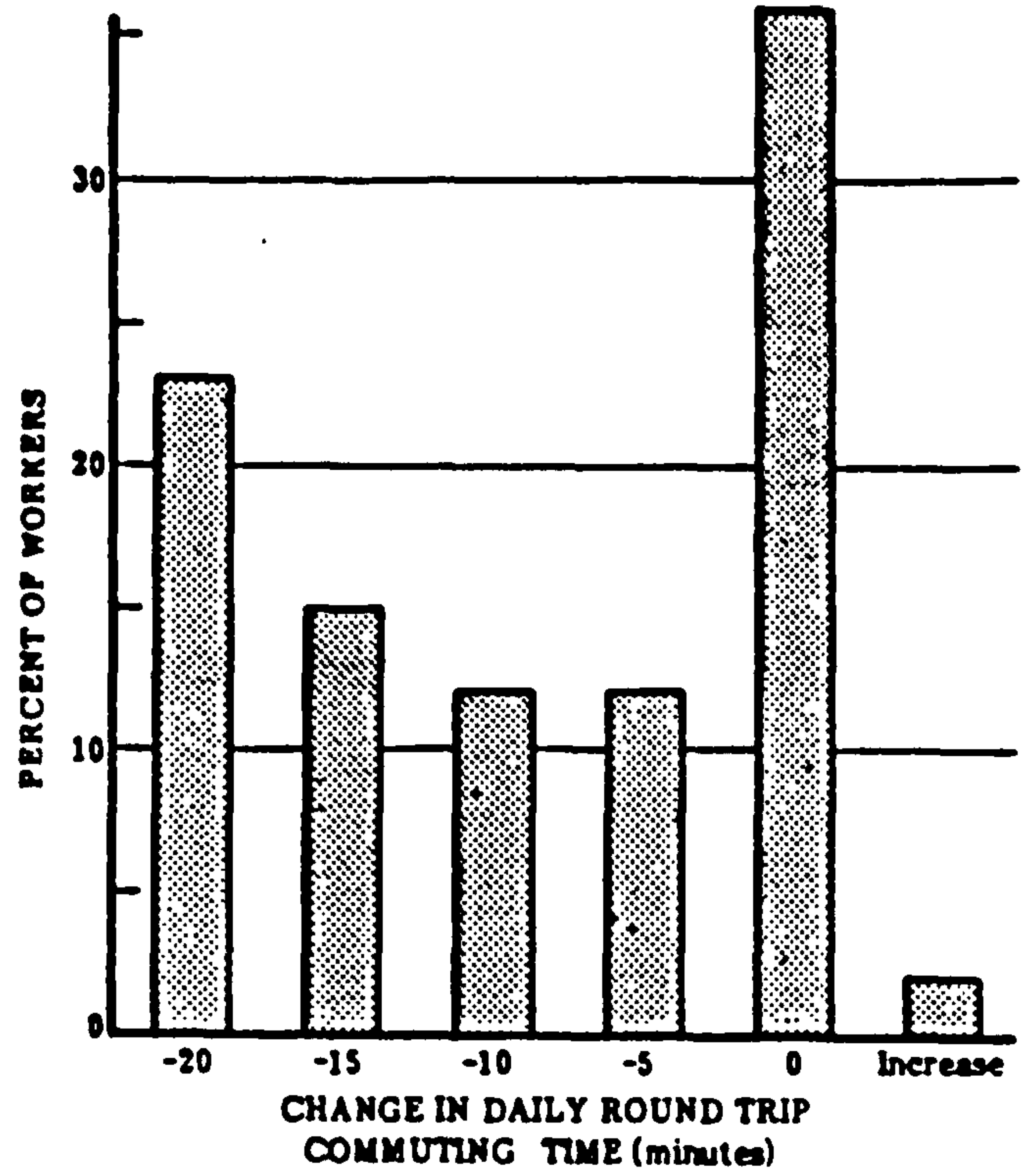


fig.4.12 Impact of EPA flexible work hours program on travel time

(Source: Alternative work schedules, TRB, 1980).

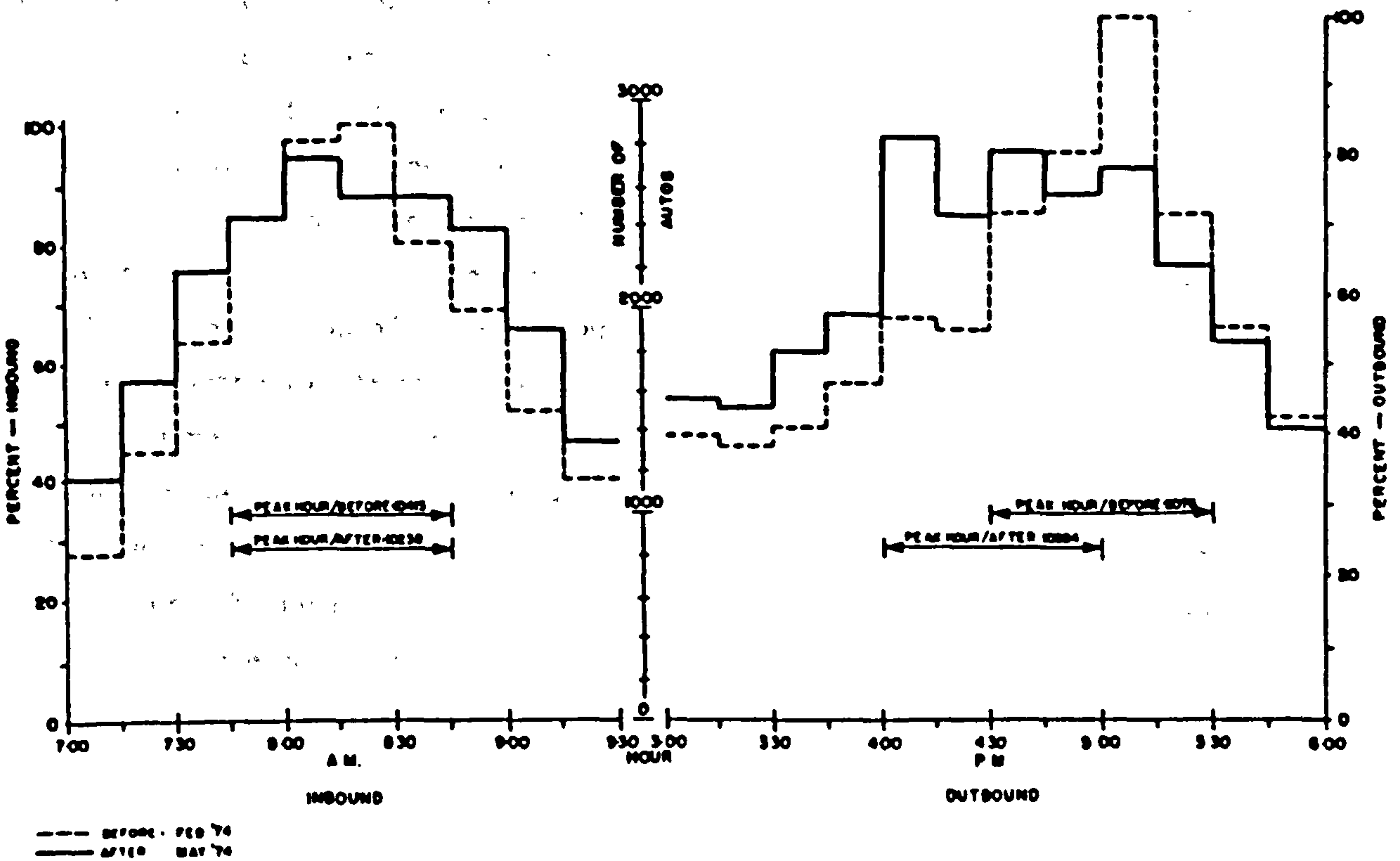


fig 4.11 Impact of variable work hours on automobile volumes at the Ottawa central area screenline.

surveys in the Ottawa study: thus the peak was flattened more significantly than indicated by the raw 15-min volume values. The peak 15-min volume, stated as a percentage of the 7:00 to 9:30 total, actually decreased by approximately 15%. The change in the peak-period traffic distribution was more significant in the afternoon. The peak 15-min period shifted to 1hr earlier, and the peak 15-min volume fell 17%, even though the 3-hr total increased by 6%. This corresponds to a 22% decrease in the peak 15-min volume, stated as a percentage of the 3:00 to 6:00 p.m. total volume.

However, perhaps the single most important consequence of variable work hours is the reduced commuter travel times. From a survey of the employees participating in the Toronto-Queen's Park variable work hours, approximately 31% said that morning and evening travel time had been shortened by an average of 11 mins, while only 3.2% stated that commuting time was longer in the morning. In Riverside, California, a staggered work hours program has estimated to result in an average reduction in commuting time by 2.5 mins. At the EPA in Washington, D.C. after the implementation of a flexible hours plan, the approximate reduction in round trip travel times was 8 min. Figure 4.12 presents that change in daily round trip commuting time; the 62% percent indicated reduction in travel times, while only 2% said travel time had increased.

As was mentioned earlier in this section another technique used to spread the work travel demand is to introduce peak hour tolls on access routes to the city.

The impact of implementing such a plan will be demonstrated here using the case of the Singapore's Area Licence Scheme. The essence of that scheme is that a special supplementary licence must be purchased and displayed in any car that is driven into a designated Restricted Zone during the morning commuting hours. A detailed analysis on the effects of that scheme is given by Watson and Holland (1978). Some of the results of this scheme are summarised below :

- The number of cars entering the Restricted Zone between 7:30 and 10:15 fell by 73%. The proportion of these cars that qualified as car pools by carrying four or more occupants rose from less than 10% to 44%.
- The volume of cars entering during the half hour before 7:30 rose by 23% as some people started their trip earlier to avoid paying the Area Licence fee.
- During the hours of restriction, speeds in the restricted zone increased by 20%, speeds on inbound radial roads increased by about 10%, while on outbound radials did not change. Speeds on the ring road fell by 20%.

More recently another road pricing scheme was introduced in Hong Kong, (Catling and Harbord, 1985) where a series of charge zones are defined and motorists are charged for each zone boundary crossing during busy times. Dawson and Brown (1985) report that implementation of this scheme has resulted in a 20-25% reduction of car trips in the congested times/places, and a 20-25% increase outside the congested times/places. In consequence traffic speeds increased significantly, particularly in the downtown areas. The impact of road pricing on travel patterns is further discussed in chapter nine, where the developed dynamic assignment model is used to assess the effects of various road pricing policies.

A variable work hours program as well as a time dependent toll scheme is, however, difficult to implement for all employees in a city. Furthermore, because of the element of individual choice which it incorporates, there can be no guarantee that it will result in a sufficient spreading of demand to justify its implementation. In many cases, therefore it would be useful to be able to predict the consequences of such programs for travel demand before implementing them. Modelling the choice of time of day to travel is a necessary requirement for an effective use of the alternative policies to manage travel demand, since it is essential in order to understand the determinants of time of day choice and to be able to predict travellers' behaviour with respect to this choice.

D'Este (1985) has developed a model to estimate the reduction in average commuter trip duration after introducing a staggered working hours policy. In his model no spatial variation in traffic speed is considered, he uses a theoretical joint probability density function of homes and workplaces, and the route over which the trip is made, is assumed to be the straight line joining the origin to destination. His work provides a general theoretical framework for analysing the impact of staggered work hours on transportation; however it is doubtful whether his model could be applied to a real situation.

Road pricing has been modelled and analysed by several authors including Smith (1979b), Lam (1988) and others. Their works can be used to evaluate the effects of a road pricing scheme, but are developed within the static equilibrium context, and therefore can predict neither the effects of a peak hour only toll, nor the possible shifting of the peak period demand.

In this section the effects of spreading peak demand over a longer period, on the level of traffic congestion were demonstrated: existing transportation facilities are more efficiently used, and travel times are substantially reduced.

#### **4.5 Summary**

This chapter has demonstrated the nature of demand and travel time peaking in transportation systems, and reviewed some studies on work trip scheduling behaviour. Different strategies which are used to alter work trip scheduling, in order to reduce the effects of peaks by spreading the demand over a longer period, were also discussed.

The peaking problem is a major feature of most cities where the majority of workers have to travel from the residential areas to the relatively small Central Business District, where the most of the employment centres are located. Thus the land use pattern, and the scheduling of working hours, dictating that most of the workers have to arrive at the same area at about the same time, gives rise to the peaking problem in urban transportation networks. The morning and evening peak hours are typical characteristics of urban commuter routes. During peak periods travel demand usually exceeds the capacity of existing transportation facilities, causing congestion and long delays which may extend far beyond the time during which demand exceeds capacity. Travel times are therefore not constant during the course of the peak period, and can be two or even more times longer than free flow travel times.

In an attempt to avoid the high travel times associated with the peaks, commuters shift their departure times in order to travel outside the rush hour. However their freedom of choice is restricted, since they have to be at work at their official work starting time. Arriving early at work may result in a shorter travel time, but is likely to involve some time wasted; arriving late however, has more severe consequences. It is the trade-off between early or late schedule delay and travel time which is crucial in the analysis of departure time choice behaviour.

Recently, the realisation of the importance of considering the time dependent nature of transportation level of service has led to the increased empirical research into the area of departure time decisions. Work scheduling decisions have been modelled within the utility maximisation context, by estimating econometric logit models of the choice of departure or arrival time with different specifications of the utility functions. These models provide estimates of the trade-off between schedule delay and travel time.

Empirical work shows that the major influences which may cause an earlier or later arrival than the particular start time are :

i) congestion avoidance : by avoiding peak congestion periods, an individual's travel time may be considerably lower

ii) Schedule delay : early or late arrivals at work may be dictated by the schedule of shared ride vehicles such as transit or carpools

iii) Service reliability : since travel times may vary from day to day, workers may plan to arrive earlier than their work start (on the average), to avoid a late arrival when travel times are longer than normal on a particular day

iv) Peak/off peak tolls and parking availability : monetary charges for parking, transit fares, or roadway facilities may vary by time of day, and therefore to induce changes in planned arrival times. Parking availability may be restricted for late arrivals as parking spaces fill up.

Further conclusions are :

- Travellers seem to be slightly more sensitive to travel time than early arrivals.
- Late arrival is much more onerous than early arrival; it seems that travellers are more sensitive to late arrival than early arrival by a factor of about 4.
- Low income and older workers tend to arrive at work earlier than the official work start time.
- Individuals employed in a professional, technical, management or administration job are less sensitive to late arrivals.

A better understanding of the factors affecting travellers departure time choice can be used to predict the extent to which travellers will reschedule their trips if congestion time changes, due, for example, to an increase in highway capacity. Expanding the capacity of existing facilities, usually involves high costs, or may be difficult due to lack of available locations for them. Furthermore, it may cause a shift in the peak without resulting in substantial reduction of travel times.

The flexibility in the departure time choice and the desire to optimise the use of existing facilities, has led to the suggestion of different strategies which attempt to spread the demand more evenly through the day. These are the staggered and flexible working hours and the peak period tolls. The effects of spreading peak demand over a longer period, on the level of traffic congestion are very promising : existing transportation facilities may not be overloaded and are more efficiently used, and therefore travel times are substantially reduced.

Conventional network analysis, as was described in chapter 3, does not consider the time dimension of the travel-to-work demand. On the other hand, in all studies of departure time decision outlined in this chapter only the demand side of the problem is taken into account, i.e. given the time varying distribution of travel times and the work scheduling constraints, the time-of-day dependent departure rates from origins are calculated. However, another important element of transport systems analysis, the effect of travel demand on the performance of transportation facilities is not considered. More recently, research was directed towards this area, and equilibrium models have been developed, which take into account the interaction between demand and network performance under time-dependent conditions. These models will be analysed in the following chapter.



# 5 a review of dynamic network analysis

## Objective

The purpose of this chapter is to review the recently developed dynamic network analysis models which take into account both the spatial and time-of-day variability of network congestion.

### 5.1 Introduction

Over the last three decades transportation network research has concentrated in the formulation, analysis and solution of models which estimate passengers' flows and travel times in congested networks. These models, already discussed in chapter 3, assume that the demand for travel can be represented by a constant (not time varying) O-D matrix, and traffic flows are uniformly distributed during the period of analysis which is an arbitrarily defined fixed time interval.

The unrealistic assumptions inherent in static assignment models which fail to represent the dynamics of traffic during the peaks, have led to the increasing interest in modelling the time-varying nature of travel demand during the peak period. The developed models were labelled by several authors as *dynamic* network models, and can be classified in two broad categories :

- The first refers to the models which estimate the time-varying flow patterns and travel times, and require that the temporal or time-varying distribution of the input O-D trip matrix be predetermined, and
- the second refers to models which require only the total O-D trip matrix, and the passengers' desired arrival times at the destinations, and predict the temporal distribution of the demand, i.e. the time dependent distribution of departure rates for each O-D pair.

A very popular dynamic assignment model classified in the first category is CONTRAM, which was developed in TRRL by Leonard *et al.* (1978). CONTRAM is a simulation program which considers the time dependent nature of demand, and uses the equilibrium principle. The movement of traffic is modelled by grouping the vehicles to form 'packets', and travel times and delays are estimated from a time-dependent queueing model. CONTRAM has been applied to several networks and provided good estimates of traffic characteristics. However CONTRAM, as well as the other models included in this category, are seriously limited by their input data requirements; a time dependent O-D matrix is difficult to obtain and even more difficult to forecast. Furthermore, these models cannot simulate the effects of variable work schedules on the temporal distribution of demand, and therefore on the time-varying flows and travel times in the network. They cannot also predict the altering of work trip scheduling and therefore the shift of off-peak demand into the peak period, after a capacity expansion in existing transportation facilities. Van Aerde and Yagar provide a review on other simulation dynamic assignment models (1988a), and also suggest a new modelling approach (1988b).

Earlier attempts on dynamic assignment modelling were made by Yagar (1971, 1976) who developed heuristics for achieving equilibrium flows under time-varying demand, approximated as uniform within sequential time slices. His method may result in an underestimation of the queueing time in the case of a queue decreasing within a time slice, and an overestimation in the case of a queue increasing in the time slice. Recently Zawack and Thompson (1987) and Hammerslag (1988) have also developed heuristic methods to solve a multiple origins single destination network, and multiple O-D pairs network, respectively.

A different class of dynamic assignment models within the first category are the ones concerned with the estimation of system optimum rather than user optimum flows. Merchant and Nemhauser (1978a,1978b), have developed such a model for a network with multiple origins and a single destination. Their model is a nonlinear and nonconvex mathematical programming problem, and was further explored by Ho (1980) and Carey (1986). However the formulation as a nonconvex program is associated with several analytical and computational problems. Carey (1987) has overcome these problems by formulating the system optimum dynamic assignment as a convex program, which he also extended to handle multiple destinations. More recently, Carey and Srinivasan (1987) have taken into account the trade-off between travel time and schedule delay, and developed a model which calculates the system optimum flows for

a multiple-origin, single-destination network requiring the aggregate demand for each O-D pair, instead of the time varying distribution of the O-D matrix.

This chapter will concentrate on the analysis of existing models that predict the time varying demand and traffic characteristics, given the total O-D trip matrix and will also review the existing empirical research on dynamic assignment. Reviews of the previous work in dynamic assignment models are also provided in the papers by Alfa (1986) and Ben-Akiva and de Palma (1987). All these models have a common structure. They consist of :

- i) a travel time model which estimates the time-varying travel times, given the time-dependent distribution of departure rates for each O-D pair and,
- ii) a dynamic demand model which reflects the decision rules adopted by travellers, and which incorporates the interrelation between demand and system's performance.

In some cases solution to the dynamic assignment problem can be achieved directly by solving a mathematical formulation expressing the interaction between travellers' decisions and system's performance. In other cases analytical solutions cannot be derived and thus an iterative procedure is used. This procedure simulates the day-to-day evolution of system conditions which depend on how users perceive and respond to congestion; the procedure is illustrated in fig 5.1. Users are assumed to review their decisions, every day,  $w$ , based on the outcome of their previous day choices. The alterations in travellers decision choices, results in a time dependent O-D trip matrix, which is used in order to calculate the network travel times. These times are then used to calculate the travellers' perceptions on trip characteristics, which will define the time dependent O-D trip matrix for the following day.

Thus at each iteration, the dynamic demand model is used (i) to update users' perceptions of the system's performance (ii) to determine the changes in trip choices and (iii) to define the time dependent O-D matrix for the next iteration step.

After a number of simulation days, it is assumed that travellers will have acquired the necessary information regarding system's performance, and therefore the system is expected to converge to a state of equilibrium, after which travellers do not change their decisions any more.

Thus the following section analyses the various travel time models used in dynamic network analysis, while in section 5.3 the different approaches used to model the interaction between demand and system's performance are discussed. Section 5.4 reviews the empirical research on this area, and section 5.5 summarises the chapter.

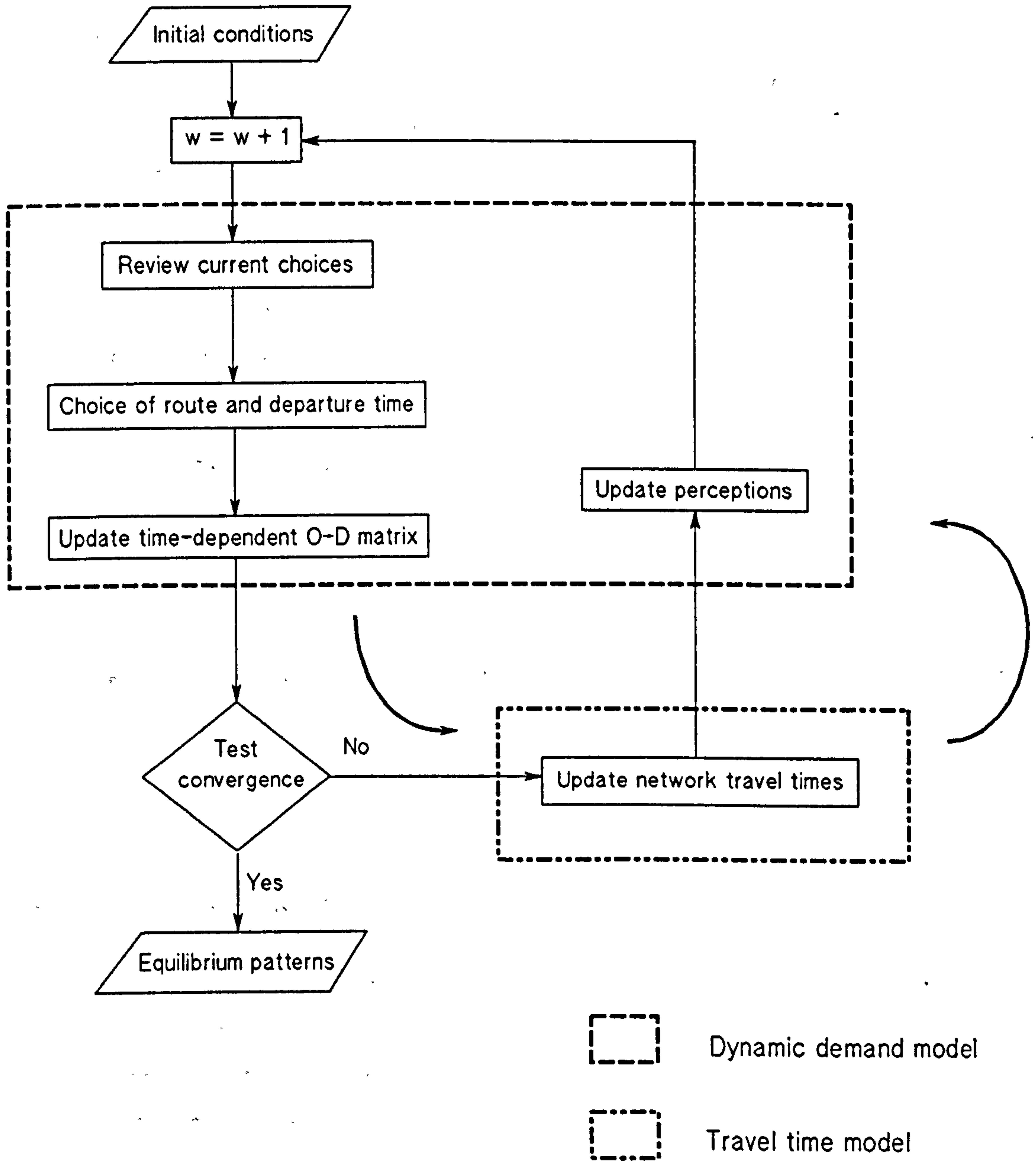


Figure 5.1: Structure of the Dynamic Model

## 5.2 Travel time models

Most of the existing research on dynamic network equilibrium dealt with the analysis of a simple network type consisting of a single O-D pair connected by a single route. More recently two extensions of this simple network have been analysed. The first corresponds to a single O-D pair connected with a number of parallel routes, and the second to a radial urban corridor represented by a highway facility which accommodates trips from several origins to a single destination. In this section the travel time models which calculate the time dependent travel times, given the temporal distribution of the demand, and which were used in the analysis of i) the single O-D pair networks, and ii) the urban transportation corridor, will be discussed.

### 5.2.1 Single O-D pair networks

Graphical representations of single O-D pair networks are given in figures 5.2a and 5.2b. Since in the multiple routes networks studied, origin is always connected to destination by parallel routes only, the travel time models used to analyse these systems are the same as for the single route network. Three different approaches were used to calculate travel times under time varying patterns of demand; they are based on i) deterministic queuing theory ii) the theory of Markov chains, and iii) traffic flow theory.

#### The Deterministic Queuing Theoretic Approach

The simple network illustrated in figure 5.2a consists of one O-D pair connected by one route with a bottleneck in between, where a queue develops in the case of congestion. Travel time from the origin O to the entrance of the bottleneck is usually assumed to be constant, whether there is a queue or not, thereby implying that the queue length is very short, as if the cars in the queue were stacked up at point B. Once a traveller departs from the bottleneck, his travel time to reach the destination D is also assumed to be constant.

In a dynamic network equilibrium, the departure rate from the origin is time dependent; thus, assuming that the bottleneck has a constant capacity,  $c$ , the waiting time can be calculated, using the *fluid approximation* approach developed by Newell (1971). This approach was employed by several researchers including Hendrickson and

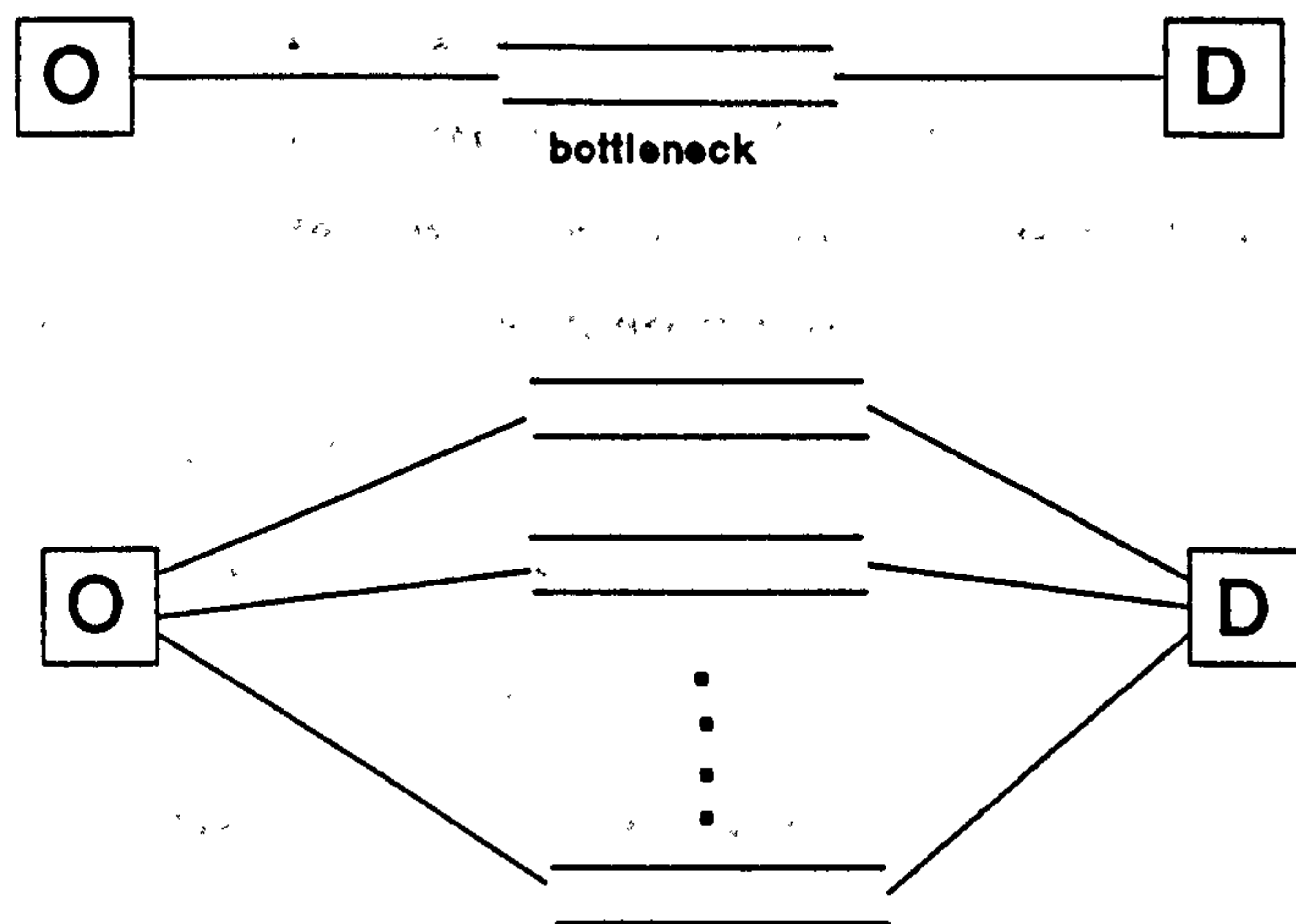


Figure 5.2: Single route and parallel routes network

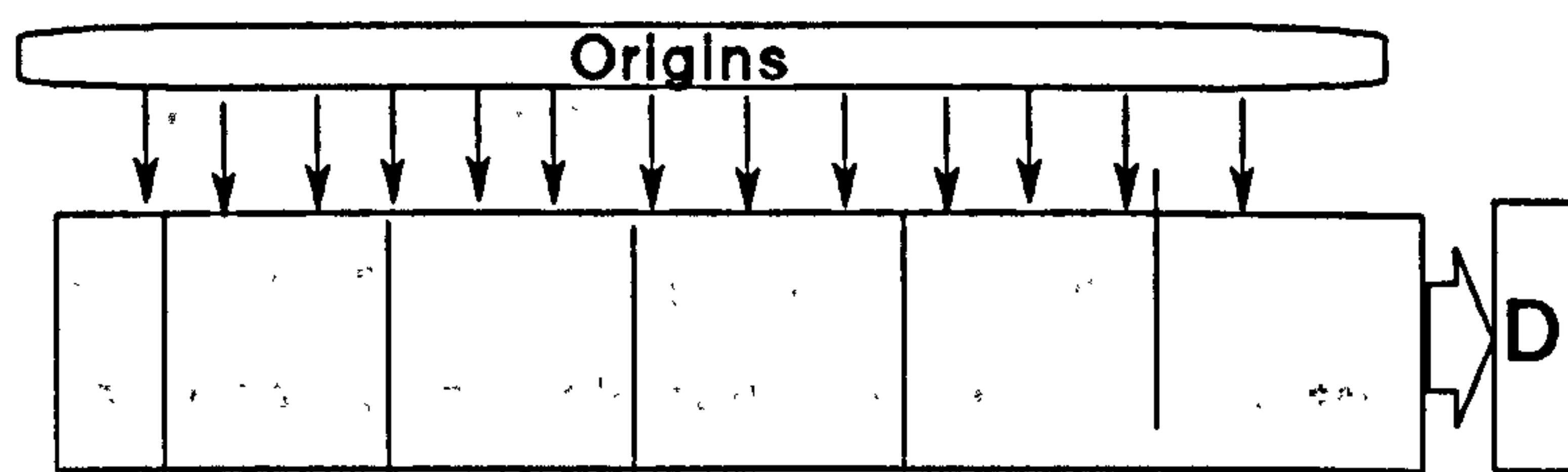


Figure 5.3 : Urban transportation corridor

Kocur (1981), Fargier (1981), Hendrickson *et al.* (1981) de Palma *et al.* (1983), Ben-Akiva *et al.* (1984, 1986, 1986a) and Kuwahara and Newell (1987).

Thus, if the number of cars in the queue at time  $t$ , is denoted by  $Q(t)$ , then the waiting time for an arrival at the tail of the queue at time  $t$ , denoted by  $tw(t)$ , is derived from the following model of a deterministic queue :

$$tw(t) = Q(t)/c \quad (5.1)$$

with no congestion  $Q(t) = 0$ , and  $tw(t) = 0$ .

The rate of change of the queue length depends on the arrival rate at the tail of the queue, and the departure rate from the bottleneck. Thus  $Q(t)$  satisfies the following differential equation based on the flow conservation relationship at the bottleneck.

$$dQ(t)/dt = \text{arrival flow} - \text{outflow} \quad (5.2)$$

Let  $AR(t)$  be the arrival rate at B. Then :

$$dQ(t)/dt = \begin{cases} 0 & \text{for } Q(t) = 0 \\ AR(t) - c & \text{for } Q(t) > 0. \end{cases} \quad (5.3)$$

de Palma *et al.* (1983) have shown that in a single bottleneck network, there can exist at most one congestion period under the dynamic equilibrium state. They further relaxed the assumption concerning the actual length of the queue, by assuming that the travel time from O to B is not constant, but depends on the number of cars in the queue. However later, Hurdle (1986) observed that there is a mathematical discrepancy in this extension of their basic model, which may lead to erroneous results.

### The Traffic Flow Theoretic Approach

Mahmassani and Herman (1984) have analysed the dynamic network equilibrium both for the case of a single and a multiple routes network. They assumed that congestion effects along a route are limited only to one roadway section of length  $l$ , thereby considering travel time as the sum of i) a flow dependent travel time on the critical section and ii) a constant term for travel along the remainder of the route. Congestion was represented using the following elementary traffic flow theoretic relationships, between density,  $K$ , flow,  $q$ , and average speed,  $v$  :

$$v = v_m \cdot [ 1 - (K/K_j) ] \quad (5.4)$$

and

$$q = K \cdot v \quad (5.5)$$



where

- $v_m$  = the maximum free flow speed  
 $K_j$  = the jam density on the roadway segment.

Let  $AR(t)$  be the arrival rate at the entrance of the critical section. Then in a given small time interval  $\Delta t$ , during which  $AR(t)$  and  $q$  can be assumed to remain constant, the net change in vehicles over the critical section is  $[AR(t)-q]\Delta t$ . Therefore the rate of change in the density within the critical section is expressed with the following differential equation :

$$dK/dt = (AR(t)-q)/l \quad (5.6)$$

Substituting equations (5.4) and (5.5) into (5.6) the following expression is derived :

$$dK/dt = (1/l).(AR(t) - v_m K + v_m K^2/K_j) \quad (5.7)$$

The above equation expresses the relationship between density and arrival rate at the critical section. Given that the departure rate pattern from the origin can be obtained from  $AR(t)$  by a simple shift of the time scale, and using equation (5.7) the relationship between speed and therefore travel time, and departure rate from the origin is obtained.

This approach was criticised by Newell (1988) as ignoring the spatial variability of traffic flow within a given road segment. For this reason Newell explored the consequences of an explicit formulation of the continuous fluid approach. In this approach the velocity of the vehicles over the links is not uniform. Point velocities are determined from a solution of a continuous flow conservation equation and an assumed speed/density relationship. His approach is theoretically sound but a lot more complicated to solve than a deterministic time dependent queuing model, or a formulation which adopts elementary traffic flow theoretic relationships.

### The Markov Chain Approach

Alfa and Minh (1979) in their dynamic assignment model have used the queuing model developed by Minh (1977). They divided the peak period in  $N$  equal time intervals and assumed that all the  $I+1$  travellers arrive at the bottleneck only at instants immediately before these intervals. Each commuter, considered separately, has the time probability  $\lambda_n$  of arriving at the bottleneck prior to interval  $(n+1)$ , given that he has not arrived at the bottleneck by interval  $n$ ,  $n=1,2,\dots,N-1$ .

Given that a traveller  $A$  has not arrived at the bottleneck by interval  $n$ , let  $C_n$  be the number of other commuters who have arrived at the bottleneck up to and including

interval  $n$ . Given any value  $k$  of the random variable  $C_n$ , the conditional probability that  $y$  other commuters arrive at the bottleneck at interval  $(n+1)$ , denoted by  $p_{y,k}^n$ , is expressed with the binomial probability :

$$p_{y,k}^n = \binom{I-k}{y} (\lambda_n)^y (1-\lambda_n)^{I-k-y} \quad \text{for } n < N-1, 0 \leq k \leq I, 0 \leq y \leq I-k \quad (5.8a)$$

$$p_{y,k}^{N-1} = \begin{cases} 0 & \text{for } 0 \leq k \leq I, y \leq I-k \\ 1 & \text{for } 0 \leq k \leq I, y = I-k \end{cases} \quad (5.8b)$$

It is also assumed that it takes  $\tau, (\tau \geq 1)$ , units of time for a commuter to go through the bottleneck, and that all commuters enter the bottleneck in their order of arrival.

Let  $L_n$  be the number of commuters both queuing up behind, and going through the bottleneck, at interval  $n$ . Let  $R_n$  be the time remaining before the commuter going through the bottleneck at interval  $n$  departs from it.

Consider the multivariate Markov chain  $\{L_n, R_n, C_n\}$  and let

$$P_{i,j,k}^n = \Pr\{L_n=i, R_n=j, C_n=k \mid L_0=0, R_0=0, C_0=0\} \quad (0 \leq j \leq \tau, 0 \leq i \leq k \leq I, 0 \leq n) \quad (5.9)$$

be the  $n$ -step transition probabilities of the chain, where by definition  $P_{0,0,0}^0=1$ . The Chapman-Kolmogorov difference equations of this chain are given by Alfa and Minh(1979).

Now, let the arrival time at the bottleneck of  $A$  be  $t$ . Then

$$\Pr\{t=1\} = \pi_1 = \lambda_0 \quad (5.10a)$$

and

$$\Pr\{t=n\} = \pi_n = (1-\lambda_0)(1-\lambda_1)\dots(1-\lambda_{n-2})\lambda_{n-1} \quad (5.10b)$$

Let  $W_i^n$ , denote the probability that an individual who decides to arrive at the bottleneck at interval  $n$ , will experience a delay of  $i$  time units. These probabilities are estimated by the following relationships :

$$W_{u \times \tau + j}^{n+1} = \sum_{y=1}^{u+1} \sum_{i=u-y+1}^{I-y} 1/(i+1) \sum_{k=y}^{I-i} P_{y,j+1,k}^n p_{i,k}^n \quad (1 \leq j < \tau, 0 \leq u < I) \quad (5.11a)$$

$$W_{u \times r}^{n+1} = \sum_{y=1}^{u+1} \sum_{i=u-y+1}^{I-y} 1/(i+1) \sum_{k=y}^{I-i} P_{y,1,k}^n p_{i,k}^n + \sum_{i=u}^I 1/(i+1) \sum_{k=0}^{I-i} P_{0,0,k}^n p_{i,k}^n \quad (0 \leq u < I) \quad (5.11b)$$

$$W_{I \times r}^{n+1} = P_{0,0,0}^n p_{I,0}^n / (I+1) \quad (5.11c)$$

where equation (5.11a) is used in the case that when a commuter arriving at the bottleneck at time  $(n+1)$ , there is another commuter who goes through the bottleneck and requires  $(j+1)$  units of time before departing from it.

### 5.2.2 Urban Transportation Corridors

The graphical representation of a radial urban corridor is given in figure 5.3. The highway facility consists of a finite number of segments with the same or different geometric characteristics where users arrive either from preceding segments or from trip origins located in surrounding areas, and travel to a single destination.

#### Traffic flow simulator

Chang *et al.* (1985) have developed a deterministic traffic simulation model, to study the traffic flow characteristics of a highway. In their model, the traffic facility is discretised into sections of uniform length  $\Delta X$ , time is discretised into small equal intervals  $\Delta T$ , and vehicles are moved in bunches, termed macroparticles, of equal size  $\Delta M$ .

For each highway section  $i$ , the following conservation equations hold:

$$K_i^{T+1} = K_i^T + \frac{1}{a\Delta X} (IN_i^{T+1} - OUT_i^{T+1} + Q_i^{T+1}) \quad (5.12)$$

where :

- $K_i^T$  = the concentration of vehicles in section  $i$  of length  $\Delta X$  during the  $T$ -th time interval in vhs/lane-mile
- $a$  = the number of lanes
- $IN_i^T$  = the number of vehicles that enter section  $i$  from the upstream section during the  $T$ -th time interval
- $OUT_i^T$  = the number of vehicles that exit section  $i$  onto the downstream section during the  $T$ -th time interval
- $Q_i^T$  = the net generation rate in section  $i$  during the  $T$ -th time interval.

It is assumed that, in each section  $i$ , the concentration and prevailing mean speed  $V_i^T$  remain constant over the interval  $[T, T+\Delta T]$ . The quantities  $K^T$  and  $V^T$  are related by an equation of the following form, which captures the interactions in the traffic system :

$$V_i^T = (V_f - V_0) \cdot (1 - K_i^T/K_0)^\alpha + V_0 \quad (5.13)$$

where

- $V_i^T$  = the mean speed in section  $i$ , during the  $T$ -th time step,
- $V_f, V_0$  = the mean free speed and the minimum speed on the facility, respectively,
- $K_0$  = the maximum or jam concentration
- $\alpha$  = a parameter

Macroparticles within a section are moved at the prevailing mean speed within that section, yielding the respective distances travelled during a particular time step. Using this speed the positions of the macroparticles at the end of the interval are then calculated. Section concentrations are subsequently updated using equation (5.12).

### The Traffic flow model

de Palma *et al.* (1984) have developed a similar model to analyse the case of an urban corridor, consisting of sections  $i, i=0,1,\dots,N$ , where the destination is section 0. They used flow conservation equations for each section, expressed as :

$$\frac{dX_i(t)}{dt} = Q_i(t) + q_{i+1}(t) - q_i(t) \quad (5.14)$$

where

- $X_i(t)$  = the number of cars in section  $i$  at time  $t$
- $Q_i(t)$  = the departure rate of vehicles originating in section  $i$  at time  $t$
- $q_i(t)$  = the flow of vehicles from section  $i$  to section  $i-1$  at time  $t$ .

Traffic conditions in a road section are assumed to be homogeneous. Therefore using equation (5.5), the flow  $q_i(t)$  can be expressed as :

$$q_i(t) = \frac{X_i(t)}{d_i/v_i(t)} \quad (5.15)$$

where

- $d_i$  = the average distance to cross section  $i$

$v_i(t)$  = the speed in section  $i$  at time  $t$ , which is assumed to be constant for a vehicle in section  $i$  at time  $t$ .

Expressions (5.14) and (5.15) and speed-flow relationships which include the geometric characteristics of each road section, are then used to define the time varying flow patterns and link travel times, given the time dependent distribution of departure rates. Origin - destination travel times are expressed as a function of the departure time and the corresponding arrival time at destination. Arrival times at each section and consequently to the destination are calculated using the following recurrence formula :

$$TA_{j|i}(t) = TA_{j+1|i}(t) + tt_{j+1}[TA_{j+1|i}(t)] \quad (5.16)$$

where

$TA_{j|i}(t)$  = the arrival time in section  $j$  for a vehicle entering section  $i$  at time  $t$

$tt_j(t)$  = the time needed to traverse section  $j$  for a vehicle entering section  $j$  at time  $t$ .

$TA_{i|i}(t) = t$

### 5.3 The Interaction between Demand and System's Performance

In the previous section, various modelling procedures were presented, which can be used to calculate the time dependent O-D travel times, given the departure rate distribution from the origin†. This distribution is a result of the aggregation of travellers' decisions regarding departure time choice, which in turn influence the time dependent distribution of travel times. This interaction between demand and system's performance is of particular interest, since it defines the steady state of the system.

Three different approaches were used to model the interrelation between demand and system's performance, and consequently to define the time dependent departure rates, flow patterns and travel times. In the first approach the problem is formulated as a Markov Chain, while the second is based on the utility maximisation principle, and the third on the bounded rationality concept.

---

† In some of the travel time models already discussed, relationships between travel time and the arrival rate distribution at a bottleneck (or a critical section in a highway), instead of the departure rate from the origin, were defined. However departure rates can be obtained from the arrival rate distribution, by a simple shift of the time scale.

### 5.3.1 The Markov Chain Formulation

Alfa and Minh (1979), who used the concept of Markov Chains to calculate traffic delays, assumed that every day, travellers review their decisions on departure time based on their experiences from the previous day. The day-to-day adjustment of commuters' decisions is modelled as an iterative procedure which can be set up as a classical Markov Chain as follows :

$$PR(\omega+1) = PR(\omega) \cdot TR(\omega) \quad (5.17)$$

where

$PR(\omega)$  the arrival times probability (row) vector  
 $= [\pi_1(\omega), \pi_2(\omega), \dots, \pi_N(\omega)]$ ,  $\pi_i(\omega)$  is the probability of arrival at time the bottleneck at time  $i$ , as defined by equation (5.10).

$TR(\omega)$  an  $N \times N$  transition matrix

The equilibrium solution is to find the  $\lim_{\omega \rightarrow \infty} PR(\omega)$ . This is obtained by assuming an initial value of  $PR(1)$  and then using (5.17) repeatedly until a steady state is reached.

Every commuter is assumed to have a *target time* at which he wishes to depart from the bottleneck in order to arrive at work on time. He also attaches costs to the travel time he experiences and to his early or late schedule delay.

In the analysis of the day-to-day decision making process, it is assumed that a traveller who arrived at the bottleneck during the time interval  $n$  on the  $\omega$ th day, would considered changing his arrival time to interval  $m$  on the  $(\omega+1)$ st day only if he thinks he could reduce his incurred cost by doing so. If he was delayed  $i$  units of time on the day he arrived at the bottleneck at interval  $n$ , and  $j$  units on the day he arrived at interval  $m$ , then the reduction associated with this particular change of arrival time will be :

$$[C(n,i) - C(m,j)]^+$$

where

$C(n,i)$  = the total cost of travel due to travel time and schedule delay for a traveller who arrived at bottleneck at interval  $n$ , and experienced a delay of  $i$  time units

$[\varphi]^+$  = is a mathematical operator such that  $[\varphi]^+ = \max[0, \varphi]$

Thus the expected reduction in cost, when a traveller changes his arrival time from interval  $n$  on the  $\omega$ th day, to interval  $m$  on the  $(\omega+1)$ -st day, denoted by  $EC_{n,m}(\omega)$ , is expressed as :

$$EC_{n,m}(\omega) = \sum_{j=0}^{I \times \tau} \sum_{i=0}^{I \times \tau} [C(n,i) - C(m,j)]^+ W_i^n(\omega) W_j^m(\omega) \quad (5.18)$$

where  $I$  travellers are assumed to pass through the bottleneck every day; it takes  $\tau$  units of time to go through it; and the probability that a traveller who arrives at the bottleneck at time interval  $n$  will be delayed for  $i$  time units, is denoted by  $W_i^n$ , and defined by equation (5.11).

In the above equation (5.18), the delay distribution used for the new decision on arrival at time  $m$ , is given for the  $\omega$ th day instead of the  $(\omega+1)$ -st as it would be expected. This is because travellers do not have the complete information about this distribution for the  $(\omega+1)$ -st day. It is therefore suggested that each traveller assumes that the other commuters do not change their arrival process for the  $(\omega+1)$ -st day, and therefore the delay distribution will remain as it was on the  $\omega$ th day.

As mentioned above, an individual's choice of arrival at the bottleneck for the  $(\omega+1)$ -st day is dependent on his expected reduction in cost. Alfa and Minh have suggested that although in practice, there are some commuters who would pick the next day arrival time to the bottleneck such that the expected reduction in cost is maximum, there are others who would avoid this time, knowing that if all arrive at the same time, the delay would be high. For this reason it is suggested that the probability that an individual would change his arrival time from  $n$  on the  $\omega$ th day, to  $m$  in the  $(\omega+1)$ -st day, is proportional to the expected reduction cost associated with this change. Upon normalising, this probability is estimated as :

$$P_{n,m}(r) = \Pr\{ t(\omega+1)=m \mid t(\omega)=n \} = \frac{EC_{n,m}(\omega)}{\sum_{m=1}^N EC_{n,m}(\omega)} \quad (5.19)$$

The transition matrix  $TR(\omega)$  of equation (5.17) is then such that  $[TR(\omega)]_{nm} = P_{n,m}(\omega)$ . The vector of probabilities  $PR(\omega)$  is then used in equation (5.10) to define the probabilities  $\lambda_n$ , which are in turn used to calculate the probabilities  $W_i^n(\omega+1)$ .

Alfa (1981) has extended the model described above to include departure time and route selection, for a single O-D pair network connected with parallel routes. Yet, there has been no proof of existence or uniqueness of solutions to any of these models.

However Alfa (1982) has shown the existence and uniqueness of solution for a slightly different stochastic model, in which he examines the decision making process by which only one traveller selects his departure time, considering that all the other travellers previously using the transportation system have decided on their departure times and will not change them under any conditions.

The model suggested by Alfa and Minh, reflects the stochastic nature of travel characteristics, and represents the dynamics of travel demand, (both within the peak period, and from day to day). However it requires a large amount of computational effort and therefore is rather impractical to apply or even to further develop so that it will handle more realistic networks. Furthermore, its assumption concerning traveller's decisions is not supported by any theory of traveller's choice behaviour.

### 5.3.2 Utility maximisation approach†

The main assumption in the utility maximisation approach is that each traveller decides on what time to depart from his origin, and which route to follow in order to reach his destination, in such a way that his total utility of travel is maximised (or disutility is minimised). Most of the models analysing dynamic network equilibrium are based on this assumption. The major differences in these models lie on the principles adopted in order to model how a traveller perceives this minimum disutility, and how it affects his decisions. Deterministic Dynamic User Equilibrium (DDUE) and Dynamic Stochastic User Equilibrium (DSUE) are the principles used to study this problem.

#### Deterministic Equilibrium

Dynamic deterministic user equilibrium is defined as the state at which *no traveller can reduce his total disutility of travel by changing his decision on route and departure time selection*. It is assumed that travellers have identical tastes, and that there are no errors involved in the perception of the actual trip characteristics; therefore travellers' perceived travel times and schedule delays are equal to the actual ones. Therefore at the equilibrium state, every traveller will have the same utility level. The DDUE approach was adopted by several researchers including Hurdle (1974), Hendrickson and Kocur (1981), Fargier (1981), Hendrickson *et al.* (1981), Mahmassani and Herman (1984), and others.

---

† In this section some of the assumptions inherent in the utility maximisation concept are relaxed. Thus, each alternative is associated with a utility which, however, is not necessarily random. The concept is used only to reflect a decision rule according to which a decision maker select the alternative which maximises his utility.



Generally it is assumed that all the travellers have the same desired arrival time at the destination,  $A$ , (which may be considered as the work start time), and attach different values of time to travel time, early schedule delay, and late schedule delay. Therefore each trip decision is associated with a particular cost or disutility of travel, expressed as :

$$C(t) = C_w \cdot TT(t) + C_e \cdot [A - t - TT(t)]^+ + C_l \cdot [t + TT(t) - A]^+ \quad (5.20)$$

where

- $C(t)$  the total cost of travel associated with departure time  $t$ .
- $TT(t)$  the travel time from origin to destination, when departing at time  $t$
- $C_w, C_e, C_l$  the value of travel time, early, and late schedule delay, respectively

Since at equilibrium no user can reduce his travel cost by unilaterally changing departure times, cost of travel should be constant for all  $t$ , in other words the change in user cost with respect to the selected departure time must be zero for any individual; this condition is expressed as :

$$dC(t)/dt = 0 \quad (5.21)$$

Furthermore,  $C(t)$  as defined in equation (5.20) is a function of travel time, which in turn, as was shown in the previous section is a function of the departure rate distribution. Therefore, by making the necessary substitutions, and then solving the differential equation (5.21), the DDUE distribution of departure rates is obtained. This distribution can then be used to calculate the time dependent travel times and flow patterns.

Hurdle (1974) did not set up the cost structure expressed by equation (5.20). He assumed that a traveller wishes to depart from his origin as late as possible without arriving at destination late. By this assumption, he implied no late arrivals at destination; hence cost of lateness is infinitely large. Hurdle did not study the case where all travellers have the same desired target time. Hendrickson and Kocur (1981) and Fargier (1981) studied the effects of staggered work hours, by introducing a general distribution of work start times. Hendrickson *et al.* (1981) have studied this problem, and also incorporated stochastic travel times, by assuming that travel time follows a uniform distribution with mean equal to the estimated travel time. Fargier also considered the work-to-home model, i.e. the evening peak problem. He proved the existence and uniqueness of the equilibrium solution for the evening peak case.

Mahmassani and Herman (1984) extended their model to analyse route and departure time decisions in the case that two parallel routes connect the origin to destination. Smith (1984) proved the existence of an equilibrium distribution of arrivals at a single bottleneck for the home to work situation, i.e the morning peak case; later Daganzo (1985) proved the uniqueness of the solution.

Simulation experiments were conducted using the models discussed above. Major conclusions from these experiments are :

- Departure rate and travel time distributions depend on the relative values of schedule delay and travel time. The derived patterns of departure time decisions are generally consistent with the observed patterns (Hendrickson *et al.*,1981)
- No increase in capacity is capable of eliminating peak period congestion with many workers starting work at the same time. Users will simply arrive later with higher capacity roads, reducing both their schedule delay and travel time, but not eliminating their queuing time completely. Policies to increase capacity will shorten the peak period and lessen delays but cannot produce free-flow conditions.

A more general network structure was analysed by Kuwahara and Newell (1987), who analysed a specific network type consisting of multiple origins connected by overlapping routes to a single destination. However, the network structure is restricted so that to allow travellers to pass through at most two bottlenecks, in order to reach their destination. The DDUE problem reduces to solving first order differential equations for schedule delay and queuing costs, which can be solved numerically. However, the proposed algorithm can only analyse networks with specific topography and thus it cannot be used for solving the dynamic assignment problem any real urban network.

Recently, Newell (1987) used a deterministic time dependent queuing model to extend existing models to the situation in which different travellers may attach different values to queuing and schedule delay. He solved this dynamic equilibrium problem for the case that no lateness is allowed, by introducing a distribution of the ratio  $C_e/C_w$  ; his model is general to allow any arbitrary distribution of this ratio. In order to allow lateness he introduced a distribution of the ratio  $C_l/C_w$  , such that  $C_l/C_w = g(C_e/C_w)$  where  $g(x) > x$ . His model is more realistic since it incorporates the variability in travellers characteristics; however it is more complex and therefore difficult to apply or to extend to a more general network. An interesting conclusion from his study of the impact of time-dependent tolls on trip decisions is that, travellers who shift their times if there were a queue are not necessarily the same people who would shift if there

were a toll. The latter person is likely to be one with a low income who values money more than time. The former person could be one with a high income and flexible work hours who does not like to waste time in queue. He concluded that charging time-dependent tolls may, therefore, be politically controversial as discriminating against the poor.

The deterministic dynamic assignment models presented above, provide analytical solutions and therefore avoid the high computational requirements associated with the Markov chain formulation. However a deterministic approach implies several restrictive assumptions regarding the travellers' characteristics, and the attributes of the alternatives. Furthermore, the existing deterministic models analyse only simple network types, and it is doubtful whether analytical solutions could be derived for more realistic network structures.

### Stochastic Equilibrium

Following the definition of stochastic user equilibrium as applied to static assignment (section 3.1), dynamic stochastic user equilibrium is defined as the state at which *no user believes he can increase his total utility of travel by unilaterally changing route or departure time*. Utility is assumed to be a random variable, reflecting the inaccuracies and distortions in travellers' perceptions' of trip characteristics. The Markovian model analysed earlier in this chapter, can therefore be characterised as a DSUE model since it treats travel time and therefore the cost of travel as a stochastic variable. In this section a different approach is discussed; travel times are calculated from deterministic models, while trip decisions are modelled using a probabilistic choice model, more specifically the logit model.

This approach was introduced by de Palma *et al.* (1983) in their study of commuters' trip decisions travelling a single route network. They used a utility function, and assumed the disutility of travel time to increase linearly at a rate of  $\alpha$  per unit of travel time, as shown in figure 5.4a. They also assumed that there is a time interval  $[t - D, t + D]$ , where  $D \geq 0$ , which defines the desired period of arrival at the destination;  $t$  denotes the centre of the period and  $D$  can be considered as a measure of work start time flexibility. The disutility of schedule delay, illustrated in figure 5.4b, is zero inside the interval  $[t - D, t + D]$ ; outside this period, schedule delay disutility of early arrivals increases linearly at a rate of  $\beta$  per unit time, and schedule delay disutility of late arrivals increases at a rate  $\beta\gamma$  per unit time, where  $\gamma > 1$ . The utility associated with a departure at time  $t$ ,  $V(t)$ , is therefore expressed as :

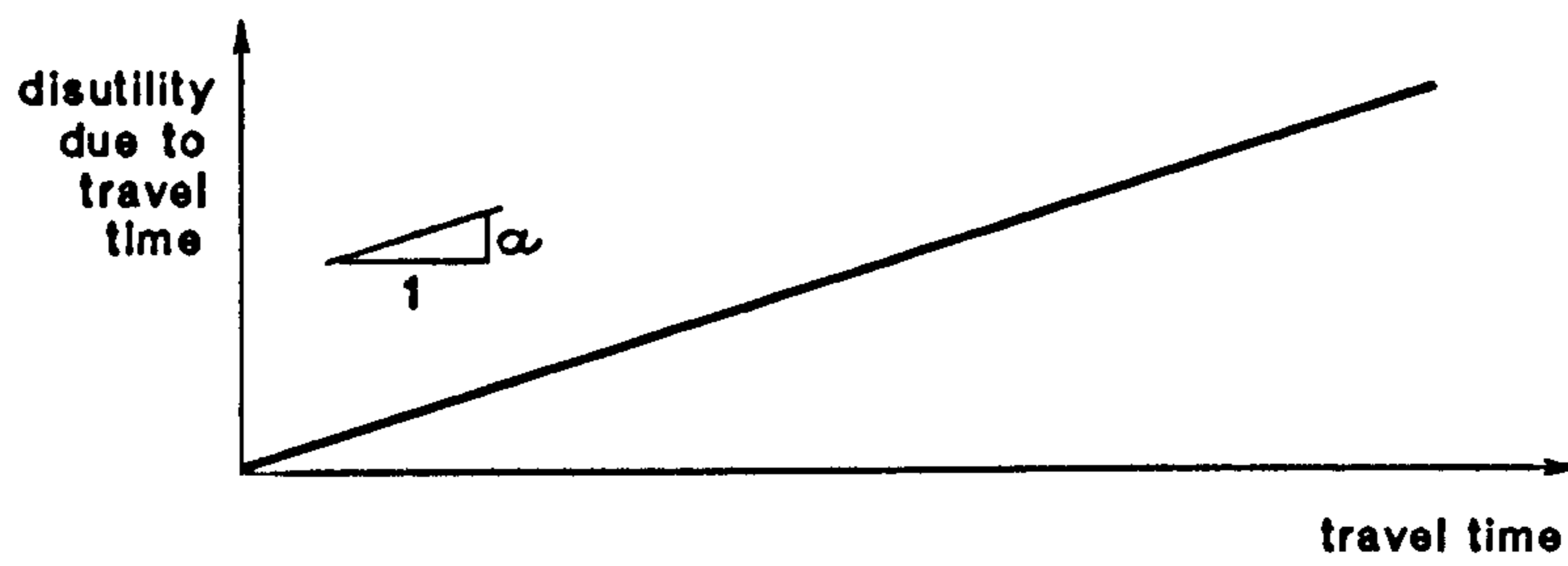
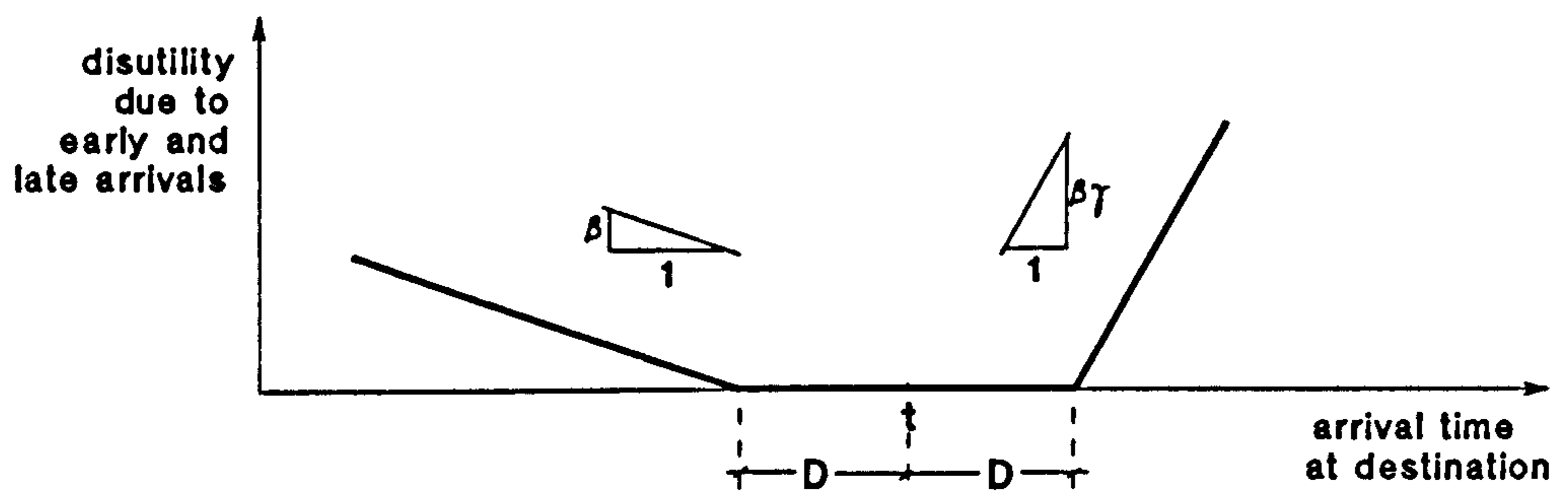


Figure 5.4: Disutility due to late or early arrivals at the destination and due to travel time.

$$V(t) = - (\alpha.TT(t) + \beta.[(t - D) - t - TT(t)]^+ + \beta\gamma.[t + TT(t) - (t + D)]^+ ) \quad (5.22)$$

Having defined the utility of travel as a function of departure time, the probability of a certain departure time being selected, is calculated using the following continuous logit model :

$$P(t) = \frac{e^{\mu V(t)}}{\int_T^{T+T_0} e^{\mu V(u)} du} \quad (5.23)$$

where the first possible departure from origin occurs at time  $T$ , and the last one at time  $T+T_0$ .

The departure rate,  $Q(t)$ , distribution is therefore obtained using the following expression :

$$Q(t) = Q \cdot P(t) \quad (5.24)$$

where  $Q$  is the total travel demand during the study period

Using the above relationships and the deterministic queuing model described in the previous section, de Palma *et al.* (1983) have formulated this DSUE problem as a Bernoulli differential equation, and derived an analytical expression of the departure rate distribution. However due to the nonlinearities involved in the analysis of more complicated networks, analytical solutions can not be derived. Thus Ben-Akiva *et al.* (1986, 1986a) who studied a single O-D pair connected with parallel routes and, de Palma *et al.* (1984) who studied an urban corridor, employed the Markovian approach suggested by Ben-Akiva *et al.* (1984), in order to derive the DSUE departure rate distributions. This approach is based on the methodology on dynamic choice modelling which was developed by de Palma and Lefevre (1983) and is discussed in section 2.7. Thus the variables involved in the problem are the same as above, with additional notation to indicate the day-to-day variability. Thus, first the number of commuters departing from the origin during the interval of time  $[t, t+\delta t]$  on day  $\omega$  is defined as :

$$x(t,\omega) = \int_t^{t+\delta t} Q(u,\omega) du \quad (5.25)$$

where  $Q(t,\omega)$  is the departure rate at time  $t$  on day  $\omega$ .

It is assumed that a fraction of individuals, denoted by  $F(t,t',\omega)\Delta\omega$ , having reviewed their decision, shift from a departure during  $[t, t+\delta t]$  to a departure during  $[t', t'+\delta t]$  during the time interval  $[\omega, \omega+\Delta\omega]$ . The rate of change of the number of individuals

departing during the interval  $[t, t+\delta t]$ , can then be expressed as the difference per unit of time between the number of individuals shifting to  $[t, t+\delta t]$  and the number of individuals shifting from  $[t, t+\delta t]$  as follows :

$$\frac{\partial x(t, \omega)}{\partial \omega} = \sum_{t' \neq t} x(t', \omega) F(t', t, \omega) - x(t, \omega) \sum_{t' \neq t} F(t, t', \omega) \quad (5.26)$$

Ben-Akiva *et al.* (1984) have modelled the transition rates among different departure times with the following simple dynamical logit model.

$$F(t', t, \omega) = F \cdot \frac{e^{\mu V(t', \omega)}}{\sum_{t''} e^{\mu V(t'', \omega)}} \quad (5.27)$$

assuming a constant transition rate out of the current decision state, denoted by  $F \dagger$ . This assumption implies that the utility of a shift to a new state is not dependent on the attributes of the current state.

Equation (5.26) will then take the form of equation (2.68) and is expressed as :

$$\frac{\partial x(t, \omega)}{\partial \omega} = F \cdot \left[ Q \frac{e^{\mu V(t', \omega)}}{\sum_{t''} e^{\mu V(t'', \omega)}} - x(t, \omega) \right] \quad (5.28)$$

Let  $\delta t \rightarrow 0$  to obtain the following continuous time limit of this dynamical system :

$$\frac{\partial Q(t, \omega)}{\partial \omega} = F \cdot \left[ Q \frac{e^{\mu V(t', \omega)}}{\int e^{\mu V(u, \omega)} du} - Q(t, \omega) \right] \quad (5.29)$$

This equation describes the evolution of the departure rate from day to day. For  $\omega \rightarrow \infty$  this dynamical system reaches the stationary state.

The framework described above, was used to analyse the case of an urban corridor, and a single O-D pair network connected with parallel routes. In the latter case, however, a nested logit was used to calculate the probabilities of route and departure time selection.

The DSUE approach was used in several simulation experiments, in order to analyse travel patterns during the morning peak period. Below, some conclusions from the

---

† In other words,  $F$  is defined as the probability that a randomly chosen individual will renew his trip decision on a given day.

experiments on the more general case of the parallel routes network, (Ben-Akiva *et al.* 1986, 1986a), are presented :

- The equilibrium distributions of departure rates are continuous but do not have continuous derivatives; the travel time distributions however, follow a smooth pattern.
- Increasing work start time flexibility causes a spreading of the departure times over a longer period, and therefore results in lower travel times.
- Decreasing the parameter  $\mu$ , i.e. assuming a higher variability in travellers' preferences, has a similar effect to that from increasing the work time flexibility. Thus the departure time distribution is more uniform which tends to reduce the level of congestion.
- Increasing capacity shortens the length of the congestion period and decreases average delay. It also results in a significant shift towards later departure times. In the case that the capacity is reduced, a shift towards earlier departure times occurs.
- The dynamic evolutions of the departure rate distributions appear to have two time scales. The first period corresponds to the major shifts among the alternative routes. During the second period significant adjustments occur in the departure time distributions while the total volume on each route remains stable.

There is no proof of existence and uniqueness of the utility maximisation equilibrium state. Ben-Akiva *et al.* (1986a) noted that the rate of convergence to a steady state is dependent on the value of the review rate  $F$ . Simulation experiments show, that, for small values of this variable convergence occurs towards a unique stationary state; however for a high value of the review rate a convergence to a stationary state is not guaranteed. Furthermore, it was found (Vythoukcas, 1988), that in the case of a general network with one O-D pair (a limiting case of the networks which will be analysed in the following chapter), the convergence to an equilibrium state is also dependent on the level of demand, as compared to the capacity of the network, and the degree of variability in travellers' preferences. High levels of demand in combination with low work start time flexibility, may result in a regular oscillatory pattern, where oscillations of the departure rate distribution take place around a possible equilibrium state.

### 5.3.3 Boundedly Rational User Equilibrium

Chang *et al.* (1985), and Mahmassani and Chang (1986), have adopted Simon's (1955) *bounded-rationality* notion and the associated *satisfaction rule* in order to model the departure time choices of commuters travelling from areas surrounding an urban radial

road, to the city centre. They considered two behaviour mechanisms in order to model the commuters' decision making process. The first is concerned with the acceptability, to a given traveller, of his most recent departure time, and the second with how an unsatisfied commuter adjusts his departure time on the following day.

The first mechanism reflects a *satisfaction* or *boundedly rational* view of commuter decisions in everyday situations. A traveller is assumed to determine his departure time on day  $\omega$ , denoted by  $D_\omega$ , based on his desired arrival time  $A$ , and his estimated travel time  $t_\omega$ , according to :

$$D_\omega = A - t_\omega \quad (5.30)$$

If  $A^*$  is his actual arrival time on day  $\omega$ , the difference  $(A-A^*)$ , i.e. his schedule delay, determines the acceptability of the actual arrival time to the user. Commuters are assumed to behave as if they had an *indifference band* of tolerable schedule delay,  $[IB^e, IB^l]$ , where  $IB^e$  and  $IB^l$  are two threshold values reflecting tolerable earliness and tolerable lateness, respectively. The tolerable amount of schedule delay is expected to depend on residential location (and the corresponding distance from destination), rules at the workplace (flexible time or strict hours), and socioeconomic characteristics of the users, and therefore can be treated as a random variable. It may not be constant over time, particularly if the system is not in a steady state, but to dynamically change in response to the user's personal experience with the facility. This indifference band yields the following satisfaction mechanism or acceptability rule :

If  $IB^l \leq (A-A^*) \leq IB^e$  then accept  $D_\omega$  and set  $D_{\omega+1} = D_\omega$  ; otherwise  $D_{\omega+1} \neq D_\omega$

Thus a commuter will maintain the same departure, as long as his previous arrival time is within his indifference band.

The second mechanism represents the departure time adjustment in the event that outcome of the latest departure time decision is unacceptable. This adjustment is determined by both the current indifference band and the user's perception of the system's travel time characteristics. Two simple models for use in the departure time readjustment process, were formulated. These are the myopic adjustment and the learning model, and are discussed below:

1. *Myopic adjustment*. Under this rule a traveller responds using the experience he acquired on the latest day. He first estimates his travel time for the  $(\omega+1)$ st day, denoted by  $t_{\omega+1}$ , as the travel time on the  $\omega$ th day, increased by a fraction of his earliness or lateness :



$$t_{\omega+1} = t_{\omega}^* + a\gamma^e SD + b\gamma^l SD \quad (5.31)$$

where

- $t_{\omega}^*$  is the actual (measured) travel time experienced on day  $\omega$
- SD denotes the actual schedule delay on day  $\omega$ , defined as  $[D_{\omega} + t_{\omega}^* - A]$ ; SD > 0 if the traveller is late, and SD < 0 if early.
- $\gamma^e$  is a binary variable such that  $\gamma^e = -1$  if the traveller arrives early on day  $\omega$ , and  $\gamma^e = 0$  otherwise
- $\gamma^l$  is a binary variable such that  $\gamma^l = 1$  if the traveller arrives late on day  $\omega$ , and  $\gamma^l = 0$  otherwise
- $a, b$  are parameters, both in the interval  $[0,1]$ , reflecting the attitudes towards earliness and lateness, respectively.

The resulting  $t_{\omega+1}$  is used to readjust the departure time on the  $(\omega+1)$ st day according to equation (5.30)

2. *Learning rule.* Under this rule a traveller is assumed to estimate his travel time using the accumulated experience acquired through repeated trials on the previous days, according to the following expression :

$$t_{\omega+1} = \sum_{u=\omega_0}^{\omega} w_u t_u^* \quad (5.32)$$

where  $w_u$ ,  $u=\omega_0, \dots, \omega$ , denotes a set of nonnegative weights attached to each day (starting with the initial day  $\omega_0$ ), with  $\sum_{u=\omega_0}^{\omega} w_u = 1$ . The relative magnitudes of the weights reflect the relative importance placed on distant versus recent experience. This learning rule was used by Horowitz (1984) to analyse the stability of static stochastic equilibrium in a two-link network. Horowitz has also suggested that recently acquired experience is likely to influence in a much greater extent the current perceptions than more distant experience does. This assumption was adopted in the bounded rationality user equilibrium approach, by formulating a special case of equation (5.32) :

$$t_{\omega+1} = (1-w_{\omega}) \cdot \left[ \sum_{u=\omega_0}^{\omega-1} \frac{1}{\omega-\omega_0} t_u^* \right] + w_{\omega} t_{\omega}^* \quad (5.33)$$

where  $w_{\omega}$  is a constant weight placed on the latest experienced travel time, such that  $0 < w_{\omega} \leq 1$ . All the days prior to the latest one are thus given a total weight  $(1-w_{\omega})$ , which is assumed to be equally distributed among all prior days.

The two behavioural mechanisms reflecting commuters' decision making process, and a dynamic travel time model can then be used in the iterative procedure which was discussed in the introduction of this chapter, and is illustrated in figure (5.1); after a number of iterations the system is expected to converge to a state of equilibrium in which all users are satisfied with their current choices, and thus do not intend to switch departure times. Bounded rational user equilibrium (BRUE), is thus achieved when all users' arrival times at destination are contained within their respective indifference bands.

Since the structure of the BRUE approach is such that, the equilibrium travel patterns can be only obtained by using simulation techniques, no mathematical proofs of the uniqueness and stability of the equilibrium solution exist. Simulation experiments in an urban transportation corridor, have shown that the system is more likely to stabilize for higher values of tolerable schedule delay. At lower values of this variable, convergence may be achieved in origins closer to their destination. The indifference band can therefore be expected to increase with distance from the destination. Furthermore, convergence to a steady state was found to be dependent on the level of travel demand. Thus, if demand is increased, a wider indifference interval may be required to reach a equilibrium.

Mahmassani and Chang (1987) have provided an analytical framework to explore the existence, uniqueness, and where applicable the stability of a BRUE in an idealised commuting system with a single bottleneck, under both identical and nonidentical work start times. They have shown that the existence of a BRUE depends on the level of demand, capacity of the bottleneck and width of the indifference band; they further concluded that a BRUE pattern when it exists, is not likely to be unique, but part of a family of possible flow patterns.

#### **5.4 Empirical research**

Travel times and schedule delays experienced by individual travellers as a result of their own decisions, depend on the transportation system's characteristics, as well as on the decisions made by a large number of other tripmakers. However, the interactions which take place in traffic systems and define the actual travel patterns are complex and difficult to model. An important component in efforts to understand the dynamics of the interaction of individual decisions in traffic systems is the observation of actual behaviour. Yet, as Mahmassani and Chang (1985) note, the collection of the necessary data in the real world presents several difficulties, including;

- the need to monitor in great detail both user decisions and the facility's time varying congestion levels over a period of at least a few weeks,
  - the high degree of experimental control required,
  - the participation of a sizeable fraction of users affecting a facility's performance, and
  - the need of sufficient resources,
- which make such an approach impractical and prohibitive.

An alternative strategy to overcome these difficulties, is to observe users' behaviour under controlled situations, using actual commuters in a simulated traffic system. This approach was suggested by Mahmassani and Chang (1985), who conducted an experiment with commuters in a hypothetical though realistic commuting situation. Their experiment is also analysed by Mahmassani *et al.* (1986) and Mahmassani and Herman (1987), and is discussed below.

An urban commuting corridor was considered, consisting of a four-lane highway (two lanes for each direction) used by residents, living adjacent to it, for their work trip to a single work destination. The corridor is subdivided into nine identical 1-mile sectors, with the common destination located at the end of the last sector. Commuter residences are located only in sectors 1 through 5, where sector 1 is the furthest outbound. Each participant was assumed to represent a group of 20 tripmakers who make identical decisions; he was assigned to only one sector, and was initially given information regarding the highway facility (free flow speed, number of lanes) and his location, as well as his work starting time at 8:00 a.m. It was explicitly stated that late arrivals are not permitted.

Participants were first asked to state their desired arrival time. Every day they had to provide their departure time, as well as their anticipated arrival times. The departure time decisions of all individuals were aggregated into a time dependent departure pattern, which was used as input to a highway traffic flow simulation model to calculate the departure-time-dependent travel time distributions. Thus on the following day each participant individually was informed about his actual travel time, before his choice on that day. The experiment, involving 100 participants, covered 24 simulation days, by the end of which the system had evolved to a stable state.

Convergence was examined through three system descriptors: departure time distributions, average schedule delay and average travel time for each sector. The overall system is considered to reach steady state when users in all sectors stop adjusting their departure time decisions. When the departure time distributions

converge in all sectors, the other two descriptors must also converge since traffic is simulated deterministically.

Departure time distributions, for all sectors, did not change after day 21. A clear geographic pattern in the evolution to the steady-state choices was apparent, with sectors closer to the destination generally reaching their steady state earlier than more distant sectors. Commuters originating from more distant sectors, appear to encounter a greater day-to-day fluctuation of the average travel time per sector. This was expected, since travel time from a given origin, is influenced by the fluctuations occurring in all the other sectors closer to the destination. The pattern of fluctuation resulted in the greater difficulty encountered by more distant tripmakers to accommodate the greater day-to-day variability of travel time. Thus sectors closer to the destination tend to reach their steady-state values earlier than more distant ones.

The steady state travel time distributions clearly exhibited peaking features; for example a 5 minutes difference in departure time from the most distant vector could result in as much as 30 minutes increase in travel time. More distant sectors experience higher peaks than closer sectors, which naturally occur earlier in the day. However, there was an unexpected observation regarding the steady state average travel time for each sector. Thus, while travel time would be expected to increase with distance from the destination, it was found that, in some cases, the average travel time experienced by users in a closer sector was greater than in a more distant one. This unpredictable outcome seems to be a result of the interaction among tripmakers in their use of the highway facility.

Regarding users' perceptions, it was observed that the ones arriving late on day  $\omega-1$  appear, on average to anticipate travel time on day  $\omega$  to be lower than on the previous day, where those arriving early on day  $\omega-1$  anticipate higher travel time on the next day. Generally, it was estimated, users can be good travel predictors only when the system has essentially stabilised. Chang and Mahmassani (1988) have used the data from this experiment to study in more detail the travel time prediction and departure time adjustment mechanism. They concluded that there is no support for the assumption that users are systematically learning about the facility's time-dependent performance, and that departure time adjustment seems to be based on the previous travel time, modified by a safety margin which is function of the commuter's most recently experienced schedule delay.

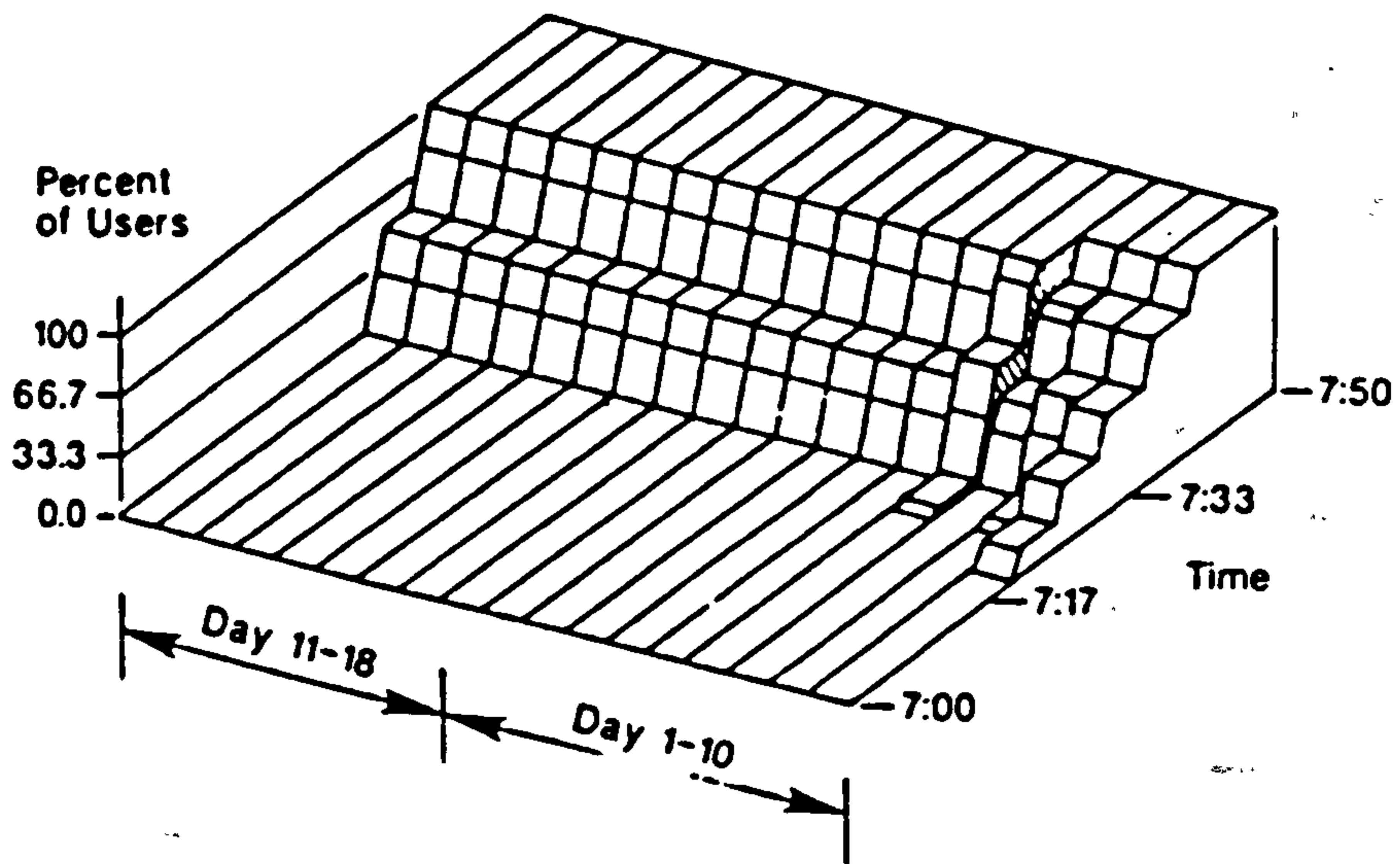


Fig. 5.5: Cumulative departure pattern evolution for sector 5 (Source: Mahmassani et al, 1986).

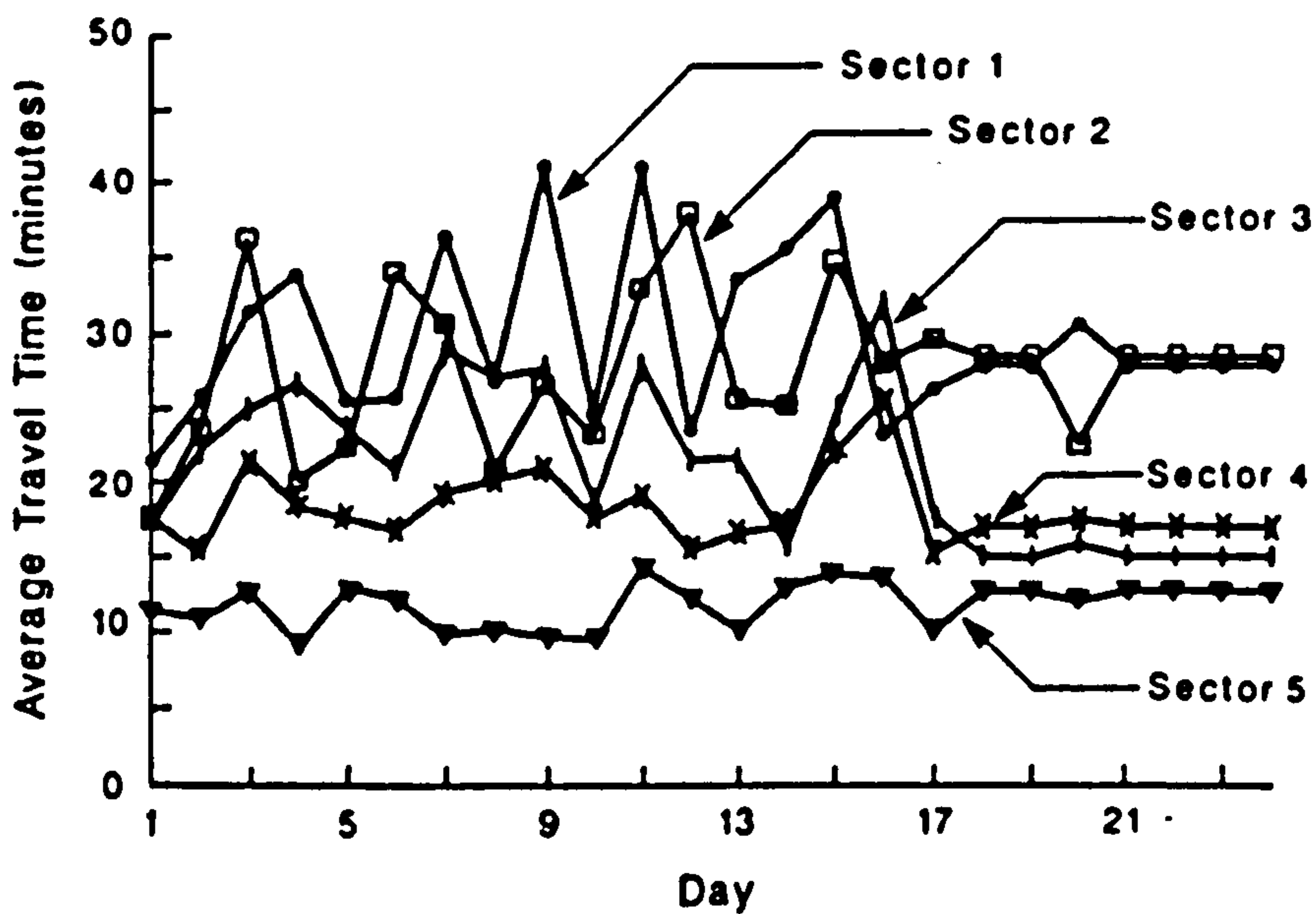


Fig. 5.6: Evolution of average travel time for each sector (Source: Mahmassani et al, 1986).

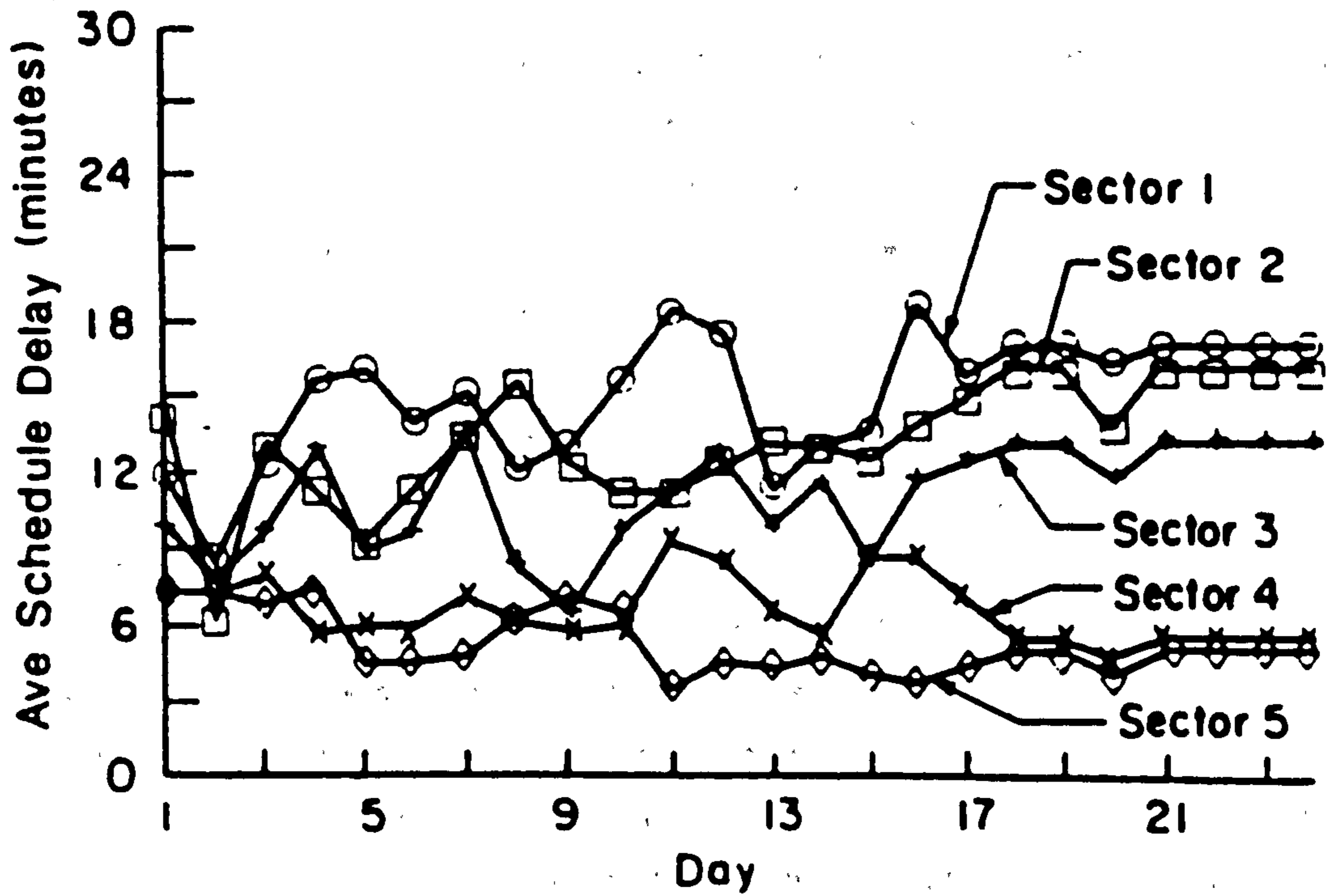


Fig 5.7: Evolution of average schedule delay for each sector (Source: Mahmassani et al, 1986).

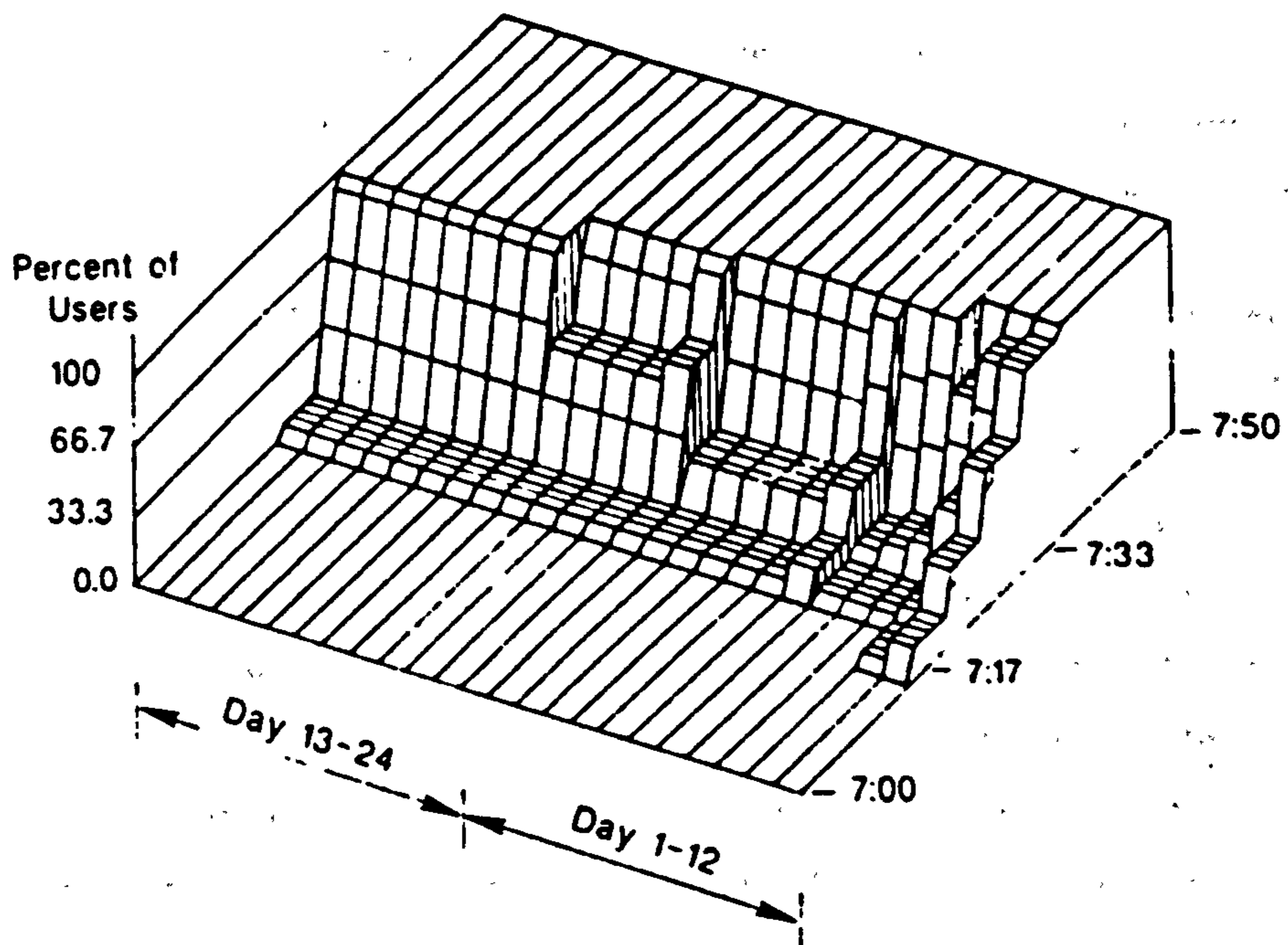


Fig 5.8: Cumulative departure pattern evolution for sector 4 (Source: Mahmassani et al, 1986).

As illustrated in figure 5.5, the departure time distribution in sector 5 has converged to a steady state at the 5th day. However, figures 5.6 and 5.7 depict that average travel time and schedule delay for that sector continue to fluctuate as long as convergence of the departure decisions has not taken place in the sectors. Thus users maintain their decisions, despite continued daily fluctuations of these performance measures. Mahmassani *et al.* (1985) used this example to support the theory of BRUE, i.e. that each user can be viewed as having an indifference band of tolerable schedule delay within which he is essentially satisfied with the outcome of his decision. They further concluded that users adjust the indifference band itself in response to their experience with the traffic system, and observed that :

- overall, users progressively increased their indifference band until it could accommodate the system's day-to-day fluctuations, and
- those who experience greater travel time fluctuations, (in this case, those in more distant sectors) ultimately had to accept substantially longer schedule delays, (particularly earliness) than those in sectors with relatively lower fluctuations, which can be more easily accommodated within narrower indifference bands.

A second experiment was conducted by Mahmassani and Tong (1986), under the same conditions as the previous one except for the informational situation, in which participants were provided with a complete profile of the system's performance on the previous day. None of the participants in the second experiments had taken part in the first one, in order to control any initial bias and learning effects. Each participant was supplied with arrival times corresponding to an array of possible departure times between 7:00 and 7:50 am in 5-min increments, from that participant's origin sector.

The first interesting result is that the system takes longer to converge under complete information than when users are provided with only their own preceding day's performance. This has happened in all the sectors of the system. Furthermore, the steady state departure distributions and all other associated performance measures resulting from the two experiments are quite distinct. Therefore, despite identical system elements and similar initial preferences of participants, two different equilibria were reached. This nonuniqueness is consistent with the results, derived by Mahmassani and Chang (1987) for an idealised situation, regarding the properties of boundedly rational user equilibrium. Overall, users are better off under the second informational situation. This is particularly true for sectors 1 and 2, where quite significant reductions of about 67% and 33% , in average trip time, were observed.

As in the first experiment, sectors in which residents encounter greater day-to-day fluctuations in system performance require a longer time to converge. However in contrast to the first experiment, in the second, sector 3 exhibited greater day-to-day fluctuations than the more distant ones. The fluctuation pattern in a given sector is a result of the complex interaction of decisions made by users in all sectors and cannot be predicted. It was concluded that there is a higher degree of interaction when users are provided with more information, which is reflected in the longer convergence times for each sector. A possible explanation is that users have greater expectations when provided with more information and may therefore have a greater willingness to experiment, which ultimately enabled them to achieve a better equilibrium state.

Using a similar example as in the first experiment Mahmassani and Tong argued that there exists a range of schedule delay that users are willing to tolerate and this range appears to increase over time, reflecting users' acceptance of progressively greater schedule delay. However, in the second experiment, the indifference band is increasing at a slower rate, which reflects users' greater expectations.

The overall conclusion from the comparison of the two experiments is that providing users with more complete information about the system's performance has induced higher expectation levels and allowed users to finally attain a better equilibrium state. However given the difficulty of learning and prediction in a system with the kind of nonlinear interactions present here, users switched with greater frequency, which resulted in longer times until convergence.

The experiments discussed above provide a valuable contribution to this area of research. However, as Mahmassani and Tong (1986) noted, when both the test situations considered in the two experiments are compared to real-world commuting systems, represent probably extreme situations. Commuters usually do not have access on information that is as comprehensive as that supplied in the second experiment. On the other hand, users might have access to more than just their own performance through word-of-mouth or media reports that they only passively receive. Therefore, real-world situations, although characterised by a certain degree of variation, tend to be somewhere between the two informational situations considered in the experiments.

Assessing the information acquired from these experiments Van Knippenberg and Van Knippenberg (1988), have suggested that a psychometric measurement method that takes account of arrival times that are preferred less than those within the indifference band but are still acceptable to people, would provide a better representation of



passengers' choice decisions. This method enables the researcher to quantify time preferences as a more continuous variable instead of a binary one as suggested by the BRUE approach. Van Knippenberg comments that it is not the indifference band that is revised. The seemingly dynamic nature of the indifference band is a consequence of the definition given in BRUE: that the indifference band is the tolerable schedule delay. They argue that the main concern of an individual is to minimise his travel time while maximising his preference for an arrival time, which is actually a two-dimensional task. This two-dimensional process can very well result in an accepted arrival time outside the indifference band but within the total acceptable interval, as long as this arrival time occurs together with a preferred travel time. An indication that travel times are important indeed, is the finding from the second experiment, that travel times for most sectors were much lower than in the first experiment. However the decision process is further complicated because the travel times are not known to the individuals and are furthermore, varying until steady state is reached.

Mahmassani *et al.* (1985) base their argument on the validity of the BRUE, by examining sector 5, where a steady state was achieved at the 5th simulation day, although in subsequent days fluctuation of travel time and schedule delay occurred.

However on examination of figures (5.6) and (5.7) it appears to be unlikely that users decide on a certain departure time solely on the schedule delay they experience, since:

- i) Users, originating from sector 5, do not accept the schedule delay of day 4, but they do accept the same amount of schedule at day 9. A possible explanation is that although the schedule delay in day 9 is higher than in day 5, travel time in the former day is much lower than the one at the latter. This implies that travellers trade-off travel time against schedule delay when they make their decisions. Furthermore, the influence of both travel time and schedule delay on commuters' decision making process, is also illustrated at day 11 when although the travel time incurred to travellers departing from sector 5 is substantially increased, travellers do not alter their decisions because schedule delay at this day is much lower.
- ii) Besides the evidence of a better equilibrium in the second experiment, individuals are willing to minimise their total disutility of travel as is demonstrated by the fact that even at the state of equilibrium in this experiment, participants were always anticipating travel time to be lower than the actual travel time.
- iii) The argument of a progressively increasing indifference band is weak, as is shown for the case of sector 4. Figure 5.8 depicts the cumulative departure time

distribution for this sector; the same state was maintained from day 7 until day 11, followed by a small shift to another state from day 13 to 16, until steady state was reached on day 17. Thus, according to the bounded rationality approach, users first accept the schedule delay on day 13, they also accept the schedule delay of day 15 which is higher, but they do not accept it at the next day i.e. they alter their departure time decisions after the 16th day although they experienced the same amount of schedule delay on the day before. The reason for that change in travellers' decisions, is probably that at day 16, travel time is much more higher than the one experienced throughout the days 13 to 15.

- iv) The piecewise linearity of the resulting cumulative departure time distribution (figs 5.5 and 5.8) is unexpected in such a complex nonlinear system.

Furthermore, the concept of tolerable schedule delay, cannot be applied in a case of a two dimensional trip decision, such as the route and departure time selection. This limitation of the BRUE approach is demonstrated with the following example.

Consider a traveller, whose indifference band is defined as [7:45 , 8:05], departs at 7:15 on day  $\omega$  , follows route  $i$ , and experiences a travel time  $t_{ti}(\omega)=25$  mins. Since his arrival time is outside his indifference band, next day he is expected to readjust his trip choices. Assuming, that he selects route  $j$ , departs at 7:20, and experiences a travel time  $t_{tj}(\omega+1)=43$  mins, his arrival time on day  $\omega+1$  will be within his indifference band, and therefore according to the BRUE approach he will be satisfied with the outcome of his latest decision. However it seems unreasonable that a traveller believes that he is better off by spending 18 mins extra in travel time instead of arriving 5 mins early.

However the concept of bounded rationality can be applied to dynamic network analysis, if the indifference band is not restricted to reflect the individual's tolerance towards early and late arrivals only. It is therefore suggested that the indifference band should be defined as a range of acceptable levels of total disutility of travel (including travel time and schedule delay), such that an individual will be satisfied, if his perceived disutility of travel is within this band. Given that each O-D pair is associated with a minimum level of disutility (i.e. when free flow conditions exist, and no schedule delay is experienced), this new form of indifference band, is equivalent to a tolerable maximum level of disutility, above which travellers will readjust their trip decisions.

The DSUE approach, which is based on the utility maximisation principle, overcomes the limiting assumptions of the boundedly rational user equilibrium, since it provides a framework which incorporates the disutilities associated with both travel time and

schedule delay. Furthermore the form of the utility function (eq. 5.22) used in DSUE, incorporates the concept of the indifference band, since this band corresponds to the interval  $[t - D, t + D]$  (fig. 5.4b).

### 5.5 Summary

Dynamic network analysis has recently received increasing attention, in an attempt to better understand traffic congestion phenomena, and to revise the inaccurate assumptions of static formulations.

Static assignment models are widely used to estimate peak period traffic flows and travel times. Peak period is arbitrarily defined as a fixed time interval, during which traffic is assumed to be uniformly distributed, and travel times to be constant. Yet, empirical observations do not support these assumptions. Flows are not uniform during the rush hour, but vary in space and time; travel times are highly dependent on the time-of-the day at which an individual decides to travel. Dynamic network analysis adds the time dimension to the static formulation, and thus provides a better representation of the features of traffic congestion.

Existing dynamic assignment models can be classified into two broad categories. The first refers to the ones which require that the time-varying distribution of the input O-D trip matrix be predetermined, while the second require only the total O-D trip matrix. A time-dependent O-D matrix is however difficult to obtain and even more difficult to forecast. Furthermore, models requiring such input data cannot evaluate the effects of demand side measures, such as flexible and staggered work hours, on the level of congestion, nor can predict the altering of work trip schedules, and the possible shift of off-peak demand into the peak period after a capacity expansion of existing facilities.

The models classified in the second category, are based on a behavioural framework, which takes into account the factors influencing travellers' choices, such as travel time, and the loss associated with an early or late arrival at destination. However, existing research on dynamic traffic assignment methods is limited to the study of simple network forms. The types analysed, include a single O-D pair connected by a single or a number of parallel routes, and an urban corridor represented by a highway facility which accommodates trips from several origins to a single destination. All the models have a common structure. They consist of a travel time model, and a demand model

which reflects travellers' decision rules, and incorporates the interaction between demand and system's performance.

Different travel models were developed to calculate the time varying travel times, given the time dependent distributions of departure rates from the origins.

The single O-D pair networks were studied using either a time dependent deterministic queuing model, or elementary traffic flow theory relationships. In the first approach the O-D pair is assumed to be connected by a route with a bottleneck in between; a queue develops at the entrance of the bottleneck when the arrival rate exceeds the fixed capacity of the bottleneck. The model analyses the evolution and dissipation of the queues using the fluid approximation approach. In the second approach congestion effects along a route are limited only to one roadway section of fixed length and limited capacity. This model provides relationships between the time dependent speed and the arrival rate at the entrance of the critical section. A different travel time model, also used, adopts the theory of Markov Chains to define the probability of experiencing certain travel times given the departure time from the origin.

In the study of urban transportation corridors, the highway facility consists of a finite number of segments with the same or different geometric characteristics where users arrive either from preceding segments or from trip origins located in surrounding areas. The two different approaches proposed, are based on traffic flow theory relationships. The first is a deterministic traffic simulation model, where vehicles are assumed to move in bunches of equal size, while the second is based on a time dependent flow conservation equation.

Three different approaches were used to model the interrelation between demand and system's performance. They are based on i) the theory of Markov Chains, ii) the Utility Maximisation principle, and iii) the concept of Bounded Rationality.

In the Markov Chain formulation, every day, travellers are assumed to review their trip decisions, based on their experiences from the previous day. Each individual attaches certain costs to travel time, and schedule delay, and is expected to change his departure time only if he thinks he could reduce his incurred cost by doing so. The probability that a traveller will change his departure time on the next day is assumed to be proportional to the expected reduction cost associated with such a change.

According to the Utility maximisation principle, travellers are expected to make their trip decisions in such a way so that they will minimise their disutility of travel. This disutility is attributed to the loss incurred due to travel time and schedule delay. In a

deterministic user equilibrium situation, no perception errors are involved, and thus all travellers experience the same disutility, which is equal to the measured disutility of travel. In the single route network case, closed form solutions can be derived. At the stochastic user equilibrium state, no traveller believes he can increase his total disutility of travel by unilaterally changing route or departure time decision. Disutility is assumed to be a random variable in order to reflect travellers' perception errors, and trip decisions are modelled using a probabilistic choice model. Closed form solutions can be derived only in the case of a single route network. Travel patterns in more complicated networks, are derived by adopting a day-to-day adjustment mechanism. The transition rates among different departure times are modelled with a simple dynamic logit model, which considers a constant transition rate out of the current state of decisions. Equilibrium is achieved on the day at which no transitions between different decision states occur.

In the Bounded Rationality approach, travellers are assumed to behave as if they had an indifference band of tolerable schedule delay, defining a time interval around the desired arrival time at the destination. If a traveller arrives at destination within this time interval he is considered to be satisfied with the outcome of his decision, and therefore he maintains the same departure time on the next day; otherwise he adjusts his departure time in an attempt to arrive within his indifference band on the next day. The departure time readjustment process is formulated using two models.

- The myopic adjustment mechanism, according to which a traveller responds using only the experience he acquired on the latest day.
- The learning rule, under which a traveller is assumed to estimate his travel time using the accumulated experience acquired through repeated trials on the previous day.

The equilibrium state is achieved when all users are satisfied with their decisions.

A method to prove the validity of dynamic assignment models, is to compare their outcomes to data collected from real world. However, due to the difficulties involved in the collection of such data, this approach is impractical and infeasible. An alternative strategy is to observe user's behaviour under controlled situations, using actual commuters in a simulated traffic system. This approach was used in order to prove the validity of the boundedly rational user equilibrium. It seems that, although BRUE provides a sound theoretical framework, its restricting assumption that the indifference band is the tolerable schedule delay is not very realistic. Travellers seem to decide on their trip choices, by trading-off the cost of travel time and schedule delay.

The models reviewed in this chapter though important and insightful (i) restrict the form of the network to one or a few bottlenecks and (ii) greatly restrict or eliminate route choice. In the following chapter a model which can handle a general network with multiple O-D pairs, connected by overlapping routes is analysed.

# **6 a dynamic stochastic assignment model for general network forms**

### **Objective**

The purpose of this chapter is to develop a dynamic stochastic user equilibrium assignment model which predicts the time varying traffic patterns and travel times in a general network.

### **6.1 Introduction**

The majority of the research effort on traffic assignment has concentrated on the development of static models (discussed in chapter 3) which however fail to represent the time varying characteristics of travel patterns and peak period phenomena (section 4.2) and ignore the time dimension on trip choice decisions (section 4.3).

On the other hand, the existing dynamic assignment models, discussed in the previous chapter, can handle only simple network forms and restrict the trip decisions to departure time choice; no route choice actually exist, with the exception of the model developed by Ben-Akiva *et al.* (1986a,b) which however can handle only an one O-D pair network connected by parallel routes.

The need to develop dynamic assignment models which can analyse general networks was pointed out by several researchers. Friesz (1985) suggests that "*..among improvements to steady-state network equilibrium models most likely to enhance their predictive capability are the needs to : include dynamic considerations....*", Ben-Akiva (1985) and Ben-Akiva and de Palma (1987) argue for the need "*..to establish the feasibility of applying dynamic equilibrium models to more realistic networks...*"

As stated in chapter 1, the main objective of this thesis is to develop a dynamic stochastic user equilibrium assignment model which can handle realistic networks.



Following the definition of static traffic assignment (section 3.1) given by Sheffi (1985), the problem known as dynamic stochastic assignment can be expressed as:

*Given :*

1. *A graph representation of the urban network*
2. *The associated link performance functions (speed density relationships or bottleneck capacities)*
3. *Travellers desired arrival times*
4. *Costs of travel time and schedule delay*
5. *Degree of variability in travellers perceptions*
6. *A total origin-destination matrix*

*Find the time dependent flow patterns and travel times on each of the links of the network.*

The problem is termed dynamic stochastic assignment since (i) the issue is how to assign the O-D matrix into the network, (ii) travellers are assumed to decide on their trip choices based on the perceived rather than measured utilities associated with the alternatives, and (iii) the assignment of the demand is performed in both space and time and the outputs of the model are time dependent patterns.

The proposed model deals with the analysis of urban road networks, and its distinctive feature is that it does not restrict the geometry of the network to certain forms. Thus the network is represented by a directed graph that includes a set of consecutively numbered nodes  $N$ , and a set of consecutively numbered links  $L$ . A link may also be denoted by its end nodes, i.e., the ordered pair  $(n_1, n_2)$  defines the link leading from node  $n_1$  to node  $n_2$ . The set of origin nodes, i.e., the nodes from which the flow enters the network, is denoted by  $R$ , and the set of destination nodes, i.e., the nodes where the flow terminates, by  $S$ . The sets  $R$  and  $S$  are such that  $R \subset N$ , and  $S \subset N$ . Furthermore, any origin node may also serve as a destination node, i.e.,  $R \cap S \neq \emptyset$ . Each O-D pair  $r-s$  is connected by a set of paths,  $k$ , denoted by  $K_{rs}$ , where  $r \in R$ , and  $s \in S$ .

In contrast to existing dynamic assignment models which handle general networks and require as an input the time dependent O-D matrix, the developed model requires only the total O-D matrix, i.e., the total number of trips between each O-D pair  $r-s$  of the network, denoted by  $Q_{rs}$ , and estimates the time varying departure rates  $Q_{k,rs}(t)$ , i.e. the number of vehicles travelling from origin  $r$  to destination  $s$ , which follow a route

$k \in K_{rs}$  and depart from  $r$  at time  $t \in [T_0, T_0+T]$ , where  $T_0$  and  $T_0+T$  is the earliest and the latest possible departure time respectively.

The model is useful for the analysis of the morning peak period. Travellers are therefore assumed to be commuters who have to be at their work place at official start times. Desired arrival times may be during the peak period and the road network may be congested. A travel time model, is developed to calculate the time dependent link travel times and from them the O-D travel times, given the pattern of departure rates,  $Q_{k,rs}(t)$ . Two different approaches are used in order to estimate the time varying traffic patterns and delays; these are based on traffic flow theory and queueing theory.

Commuters may have the choice between an on time arrival with a long travel time and a late or an early arrival with a shorter travel time. The utility maximisation decision rule is assumed to describe the mechanisms used by the traveller to process the available information and arrive at a choice of a route and a departure time. Travellers may have different perceptions of the utility derived from the same alternative, and therefore the traffic patterns resulting from the assignment procedure depend on the degree of variability of travellers' perceptions. Furthermore, their choices are assumed to result from the trade-off between travel time and schedule delay, and to be the outcome of a two-stage decision making process : (i) choice of departure time and (ii) choice of using a route  $k$  conditional to the choice of departing at  $t$ . Therefore, given the level of utility,  $V_{k,rs}(t)$ , associated with each alternative combination of route,  $k$ , and departure time,  $t$ , a demand model having the form of the nested logit model can be used to define the probability of selecting each alternative; the latter probabilities can be then used to define the departure rates  $Q_{k,rs}(t)$ . It should be noted that since the characteristics of the available alternatives change over time, the choice set considered by the decision makers may not be constant, but time dependent.

At the state of equilibrium no traveller believes that he can increase his total utility of travel by unilaterally changing route or departure time. A flow pattern of travellers will therefore represent an equilibrium if it is consistent with both demand and network performance relationships, i.e. will be the pattern that satisfies both the demand and the travel time model.

However, as will be demonstrated in the following sections, due to the complexity of the problem, equilibrium solutions can not be achieved analytically. Thus a dynamic framework is developed in order to represent the interaction between users and network performance. This framework models the evolution of the traffic patterns

from day to day and is derived from a Markovian model. Thus, the model is dynamic in two senses. First, it provides the time varying characteristics of demand and network performance within the peak period i.e. non-uniform distributions. Second it describes the evolution of traffic patterns from day to day. The steady state solution of this dynamic model is the stochastic equilibrium pattern.

The presentation of the model proceeds as follows. In the following section the time dependent route choice set is defined. Section 6.3 presents a travel time model which considers the departure rates  $Q_{k,rs}(t) \forall k,rs,t$  as given and determines the time dependent traffic patterns and O-D travel times. Section 6.4 deals with the travel demand model and predicts the departure rates  $Q_{k,rs}(t)$  given the O-D travel times. Section 6.5 describes the interaction between demand and network performance in a dynamic framework, while section 6.6 presents the structure of the simulation algorithm of the model. Section 6.7 summarises the chapter.

## 6.2 The dynamic route choice set

This section addresses the issue of the choice set definition, i.e. the set of feasible alternatives, considered by the travellers.

As mentioned in the introduction an individual will first decide on what time  $t \in [T_0, T_0+T]$  to depart, and then which route  $k \in K_{rs}$  to follow conditional to the choice of  $t$ . In the formulation of the model, the interval  $[T_0, T_0+T]$  is chosen to be large enough such that it does not affect individuals' choices. It is the set of alternative routes which needs a more thorough examination and will be examined in this section. However, before proceeding to the analysis of this set, the notation related to the representation of the routes connecting the O-D pairs of the network, is given.

In the model formulation, each route  $k \in K_{rs}$ , connecting the pair  $r-s$ , is defined as an *ordered chain* in the terminology of Ford and Fulkerson (1962), i.e. as a sequence of links :

$$(n_1, n_2), (n_2, n_3), \dots, (n_{\nu-1}, n_{\nu})$$

where  $n_1, n_2, \dots, n_{\nu}$  are distinct nodes,  $n_1 = r$  and  $n_{\nu} = s$ .

In the analysis that follows in this chapter, a path  $k \in K_{rs}$  will be defined as an *ordered set*, denoted by  $L_{k,rs}$ . This term is used as an extension of the mathematical expression 'ordered pair'. An ordered set in the context of directed networks is defined here as a set that includes links which constitute an ordered chain. Each element  $i$  of an ordered set  $L_{k,rs} = \{i_1, i_2, i_3, \dots, i_x\}$  is associated with a variable  $O_k(i)$ , which defines the order

of  $i$  within the ordered set. Thus in  $L_{k,rs}$ ,  $O_k(i_1)=1$ ,  $O_k(i_2)=2$ , ...,  $O_k(i_x)=x$ ,  $i_1$  is connected to the origin  $r$ , and  $i_x$  is connected to the destination  $s$ .  $L_{k,rs}$  will be equal to another ordered set  $L_{m,rs} = \{j_1, j_2, j_3, \dots, j_y\}$  only if  $x=y$ , and  $i_1 = j_1$ ,  $i_2 = j_2$ ,  $i_3 = j_3$ , ...,  $i_x = j_y$ .

### 6.2.1 The set of reasonable paths

The number of possible paths, which connect an origin to a destination is generally very large. It is reasonable to assume that travellers are only able to compare a small number of alternative routes, and therefore do not take into consideration the whole set of alternative routes connecting their origin to their destination. Instead they consider a subset of this set, namely the choice set which includes the routes which travellers estimate as being reasonable options.

It is assumed that an individual travelling between an O-D pair, considers a path as reasonable if it includes only links that take him further away from the origin and closer to the destination. This definition coincides with the conditions put forward by Dial (1971) in his development of a stochastic network loading algorithm based on the logit formulation, (section 3.5). However, in this study, since the time varying characteristics of travel demand are analysed, Dial's assumptions are developed and extended to reflect the time dependent traffic patterns which take place during peak periods. Thus, since link travel times do not remain constant during the peak period, it is reasonable to assume that the choice set is a dynamic variable, i.e., it does not remain the same during the peak period, but is a function of the departure time. In the remainder of this study, the dynamic choice set will be denoted by  $K_{rs}(t)$ , and will express the set of reasonable routes considered by an individual travelling between  $r$  and  $s$ , and who decided to depart from  $r$  at time  $t$ . Furthermore, the set which includes all the paths which for at least one departure time  $t \in [T_0, T_0+T]$  were considered as a reasonable choice, is denoted by  $K_{rs}$ :

$$K_{rs} = \{ k \mid k \in K_{rs}(t), \forall t \in [T_0, T_0+T] \} \quad (6.1)$$

Moreover, for reasons of computational efficiency the criteria for reasonable paths can be redefined as : A path will be considered reasonable if it includes only links that do not take the traveller back towards the origin.

For each departure time, such links can be defined by associating a departure time dependent label to each node of the network. This label, denoted by  $r(i,t)$ , is equal to

the travel time along the minimum travel time path, from origin  $r$  to node  $i$ , when departing at time  $t$ . Thus, for a traveller departing at  $t$ , a reasonable path will include only links  $i \rightarrow j$  such that  $r(i,t) \leq r(j,t)$

The dynamic nature of the route choice set can be demonstrated with the example depicted in figure 6.1. The values of  $r(.,.)$  for all nodes, are given for four different departure times  $t_1$  (before peak),  $t_2$  and  $t_3$  (during the peak), and  $t_4$  (after peak). The dashed lines represent the reasonable paths for each departure time.

It should be noted that the procedure for defining the reasonable routes requires the use of an algorithm to find for each departure time from an origin node, the shortest path to all the other nodes of the network, given the time varying links travel times. In the remainder of this section such an algorithm is presented.

#### 6.2.1.1 Dynamic shortest path algorithm

Shortest path algorithms are essential procedures for the analysis of transportation networks. Such algorithms, reviewed by Van Vuren and Jansen (1988), have been studied over the last three decades, and analyse networks with constant link travel times. The algorithm developed in this subsection can be used to find shortest path in a network with time varying link travel times and is therefore necessary in order to define the route choice set in the DSUE problem. It is an extension of the Moore-Pape's algorithm, also known as the label correcting algorithm, which as Sheffi (1985) argues, is a very efficient algorithm for use in static assignment. The aim was not to develop the most efficient algorithm, but to develop a procedure which can give accurate predictions.

The algorithm finds the shortest paths at a given departure time from a given origin (root) node to all other nodes in a network where the link travel times are time varying. It is an iterative procedure and examines all the nodes of the network. At each iteration one node is examined and the algorithm defines the currently best route from the origin to the node. The algorithm terminates when no better route can be found from the origin to any node of the network.

The network is represented by a list of links identified by their end nodes. A time dependent travel time distribution is associated with each link  $i \rightarrow j$ . In other words the travel time,  $tt_{ij}(t)$ , needed to traverse each link  $ij$  is given as a function of the time,  $t$ ,

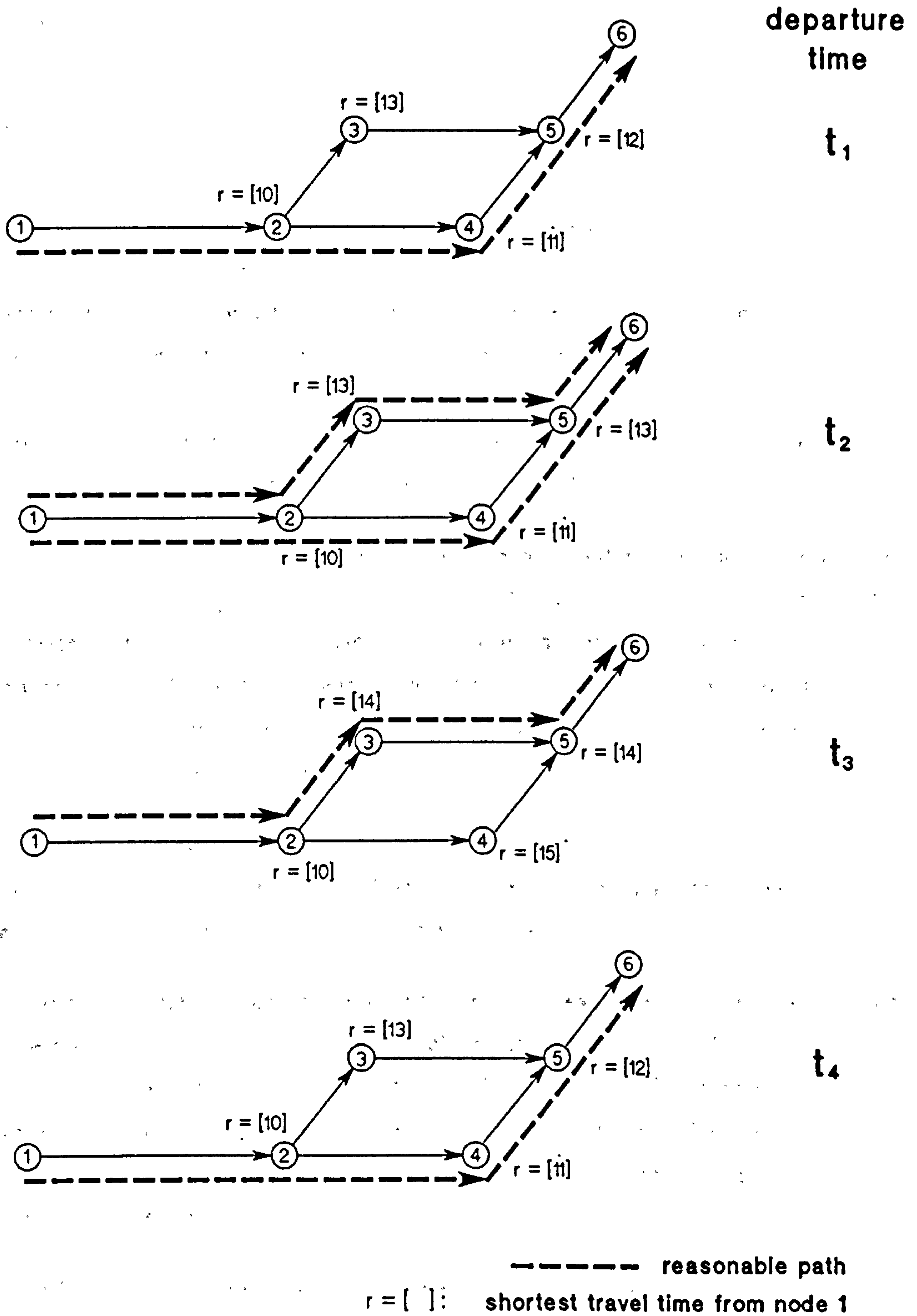


Figure 6.1 : Reasonable routes as a function of the departure time

of entering that link. With regards to the variation of the link travel times through time it is assumed that

$$\frac{d \text{tt}_i(t)}{dt} > -1$$

In addition, given the departure time,  $t_r$ , from the root, each node is associated with two other pieces of information.

(i) the current label of this node,  $l_i(t_r)$ , which defines the travel time from the root node to node  $i$  along the (current) shortest path, when departing from the root node at time  $t_r$ .

(ii) its current predecessor node,  $p_i(t_r)$ , which is the node just preceding node  $i$  along the current shortest path, when departing from the root node at time  $t_r$ .

Labels and predecessor nodes are kept in lists which are updated at each iteration in order to define the current shortest path. Given the list of the predecessor nodes, the currently shortest path can be traced backwards from any node to the root node. The algorithm examines all the nodes at least once. An additional list, called the sequence list, is also required and includes the nodes which have yet to be examined as well as the ones requiring further examination.

The algorithm starts by setting:

- the labels of all the nodes to infinity (practically to a very large number), and
- all the predecessor nodes (in the predecessor list) to zero.

Given the departure time  $t_r$  from the root node  $r$ , the next step is to place the root node  $r$  on the sequence list with label  $l_r = t_r$ . Each iteration starts with the selection of a node,  $i$ , from the sequence list for examination. Thus at the initial iteration, the root node is examined, since it is the only one in the sequence list. The nodes that are examined in the subsequent iterations are the ones that can be reached from  $i$  by traversing only a single link.

If in the iteration at which link  $i \rightarrow j$  is examined, the shortest path to  $j$  (through  $i$ ) is shorter than the previous shortest path, i.e. when

$$l_i + \text{tt}_{ij}(l_i) < l_j \tag{6.2}$$

then the current shortest path from the root node  $r$  to  $j$  can be improved by going through  $i$ . It should be noted that since the link travel times are time varying, the link

travel time, used in the equation (6.2) is the one associated with  $l_i$  as entry time on the  $ij$ .

If the inequality (6.2) is satisfied, then

- The label list is updated by setting

$$l_j = l_i + tt_{ij}(l_i) \quad (6.3)$$

- the predecessor list is updated by setting

$$p_j = i \quad (6.4)$$

- the sequence list is updated by adding node  $j$  to it. This update in the sequence list is done because the change in the shortest route to  $j$  may affect the nodes that can be reached from  $j$ .

Once all the nodes  $j$  that can be reached from  $i$  are tested, the examination of node  $i$  is complete and this node is deleted from the sequence list. The algorithm proceeds by examining the next node in the sequence list, and terminates when there is no node in the sequence list. In the following paragraphs, a simple example is presented to demonstrate the validity of this algorithm.

Thus consider the simple network illustrated in figure 6.2 . This network includes one O-D pair, and five links. Table 6.1 provides the time varying link travel times, i.e., the time needed in order to traverse a link depending on the time of entering that link. The algorithm will be used to find the shortest route from node 1 to node 4, for a journey that starts at the time  $t_r=1$ . Table 6.2 provides the contents of the label list , predecessor list and sequence list for each iteration.

First the labels of all the nodes take the value  $\infty$  and all the predecessor the value of 0; then the root node 1 is placed on the sequence list. Node 1 is the first one to be examined since it is the only one in the sequence list. The entry time to this node is set to  $t_r=1$ , and thus its label  $l_1=t_r=1$ . At iterations 1,2 and 3, the three nodes that can be reached from node 1, are examined. Assuming that node 2 is examined first, the label of node 2 will be changed since  $l_1 + tt_{12}(l_1) = 1 + tt_{12}(1) = 3 < l_2 = \infty$ . Thus  $l_2$  will take the value 3. The predecessor of node 2 will be node 1, (i.e.,  $p_2=1$ ) and the entry time of the predecessor node will be node 1 . Node 2 is placed in the sequence list, and node 1 remains at the sequence list since nodes 3 and 4 are yet to be examined at iterations 2 and 3 respectively. At iteration 3, node 1 is erased from the sequence list,



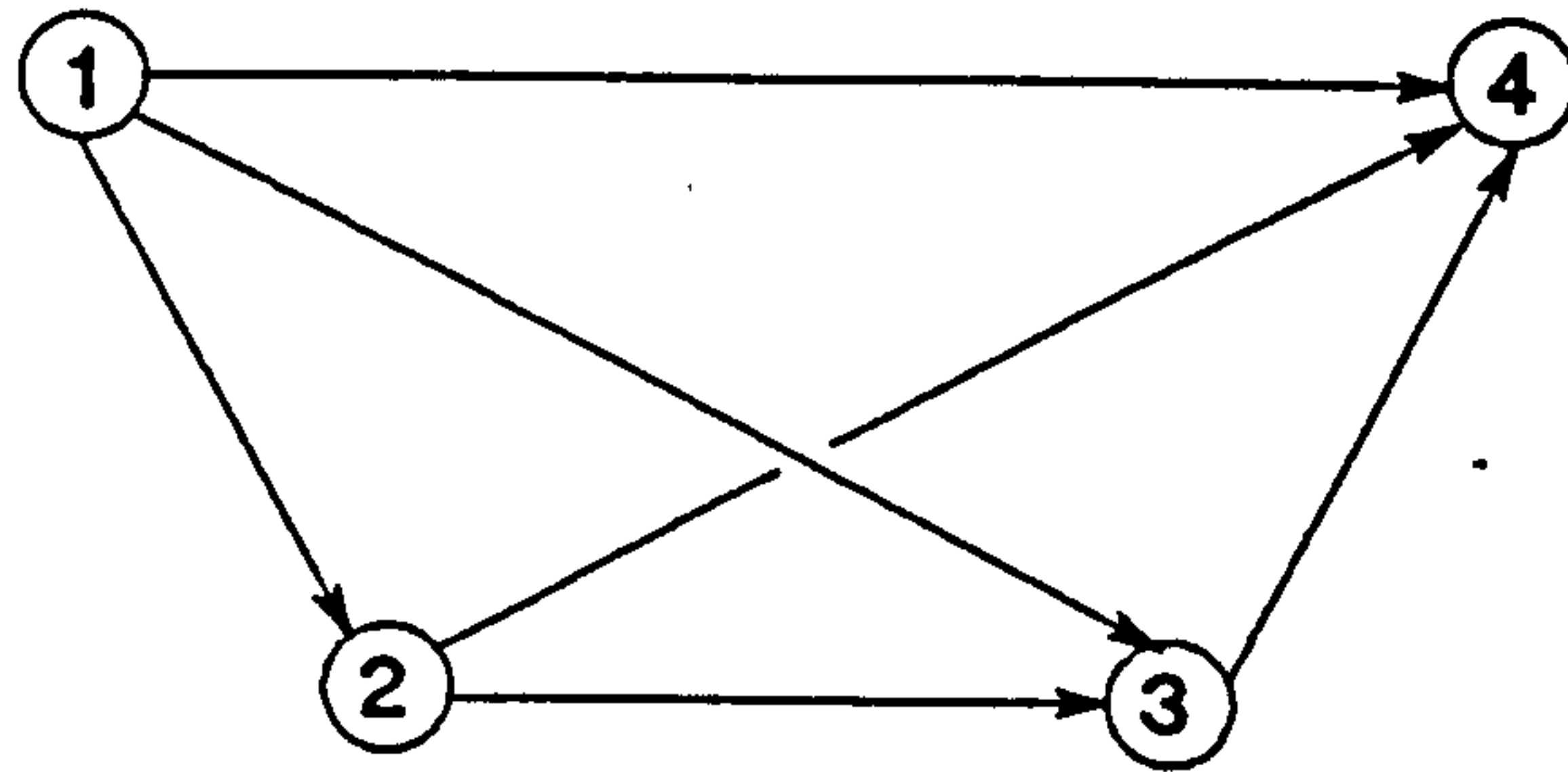


Figure 6.2 : Example network

---

entry time	travel time					
	1→2	1→3	link 1→4	2→3	2→4	3→4
1	2	4	5	1	2	1
2	3	5	6	1	3	1
3	2	6	6	1	3	1
4	2	4	5	3	3	1
5	2	4	5	2	4	3

---

Table 6.1: Time varying link travel times for the example network

---

iteration	Link tested		Label List				Predecessor List†				Sequence list
	entry time	link	Node 1	Node 2	Node 3	Node 4	Node 1	Node 2	Node 3	Node 4	
initialise	-	-	1	$\infty$	$\infty$	$\infty$	0(-)	0(-)	0(-)	0(-)	1
1	1	1→2	1	3	$\infty$	$\infty$	0(-)	1(1)	0(-)	0(-)	1,2
2	1	1→3	1	3	5	$\infty$	0(-)	1(1)	1(1)	0(-)	1,2,3
3	1	1→4	1	3	5	6	0(-)	1(1)	1(1)	1(1)	2,3,4
4	3	2→3	1	3	4	6	0(-)	1(1)	2(3)	1(1)	2,3,4
5	3	2→4	1	3	4	6	0(-)	1(1)	2(3)	1(1)	3,4
6	4	3→4	1	3	4	5	0(-)	1(1)	2(3)	3(4)	4
7	-	-	1	3	4	5	0(-)	1(1)	2(3)	3(4)	

† The figures in brackets express the entry time on each node in the predecessor list.

Table 6.2 : Contents of the label list, predecessor list and sequence list for each iteration

and at iteration 4 node 2 is examined since it is the first in the sequence list formulated at iteration 3. At this iteration the label of node 3 is changed since  $l_2 + t_{23}(l_2) = 3 + 1 < l_3 = 5$ . Therefore now  $l_3 = 4$  and  $p_3 = 2$ . At the fifth iteration node 4 is re-examined and at the sixth iteration  $l_4$  changes from 6 to 5 and  $p_4$  from 1 to 3. The algorithm terminates at iteration 7 since no link emanates from node 4 which is, therefore, removed from the sequence list.

The shortest path from node 1 to node 4 when departing at time  $t_r = 1$ , can be traced backward from node 4 by using the predecessor list.

### 6.3 The Travel time model

In this section two different models are developed to estimate the O-D travel times as a function of the departure time and the route selected, given the time varying distributions of the O-D travel demand.

- The first is based on traffic flow theory and estimates the evolution over time of the traffic flows in a general network. The time varying traffic flow patterns and speed - density relationships are then used in order to calculate the time dependent link, and from them the O-D travel times.
- The second model is based on queueing theory and determines waiting times as a function of the time dependent queue lengths developed during the peak period.

Both models are expressed by mathematical formulations which however cannot be solved analytically. Thus, they provide the framework of two *macroscopic simulation* models since they represent continuous flow profiles and queues as opposed to modelling individual vehicles or groups of vehicles. The models are based on existing traffic flow and queueing theory formulations which are used in the analysis of simple networks, and were presented in chapter 5. In this section the existing formulations are further developed so that they can be applied to any general network. The development is based on Fan's (1976) and Chang's (1977) time dependent analysis of queues in computer communication networks.

### 6.3.1 The traffic flow model

The traffic flow passing a fixed point A, of a link  $i$  at time  $t$ , can be considered as the aggregation of different flow components, each of them corresponding to the different paths, (connecting the O-D pairs of the network), which include link  $i$ .

Figure 6.3 illustrates this assumption. The flow  $q_j(t)$  which passes the fixed point A on link  $j$  at time  $t$ , consists of an aggregation of the flows that follow routes  $k_1, k_2, k_3$ , and  $k_4$ ; the traffic flow at time  $t$ , corresponding to each route  $k$ , denoted by  $q_j^k(t)$ , can then be expressed as :

$$q_j^k(t) = q_j(t) \cdot p_{j,k}(t) \quad (6.5)$$

where  $p_{j,k}(t)$  represents the proportion of the aggregate flow  $q_j(t)$  that follows route  $k$  at time  $t$ .

Furthermore, two other assumptions are included, in the analysis of the time varying traffic patterns. These were suggested by de Palma *et al.* (1984) in their analysis of an urban transportation corridor, (which was discussed in section 5.2.2), and by Ben-Akiva (1985); the latter put forward these assumptions in the development of a theoretical framework for dynamic network equilibrium models. Thus :

(i) *The traffic conditions within a link are assumed to be homogeneous.* In other words vehicles are assumed to be uniformly distributed over the length of a link. Therefore the flow  $q_j(t)$  passing a fixed point of a link  $j$  at time  $t$ , is equal to the flow leaving that link (i.e. the outflow) at time  $t$ , denoted by  $LOUT_j(t)$ ,

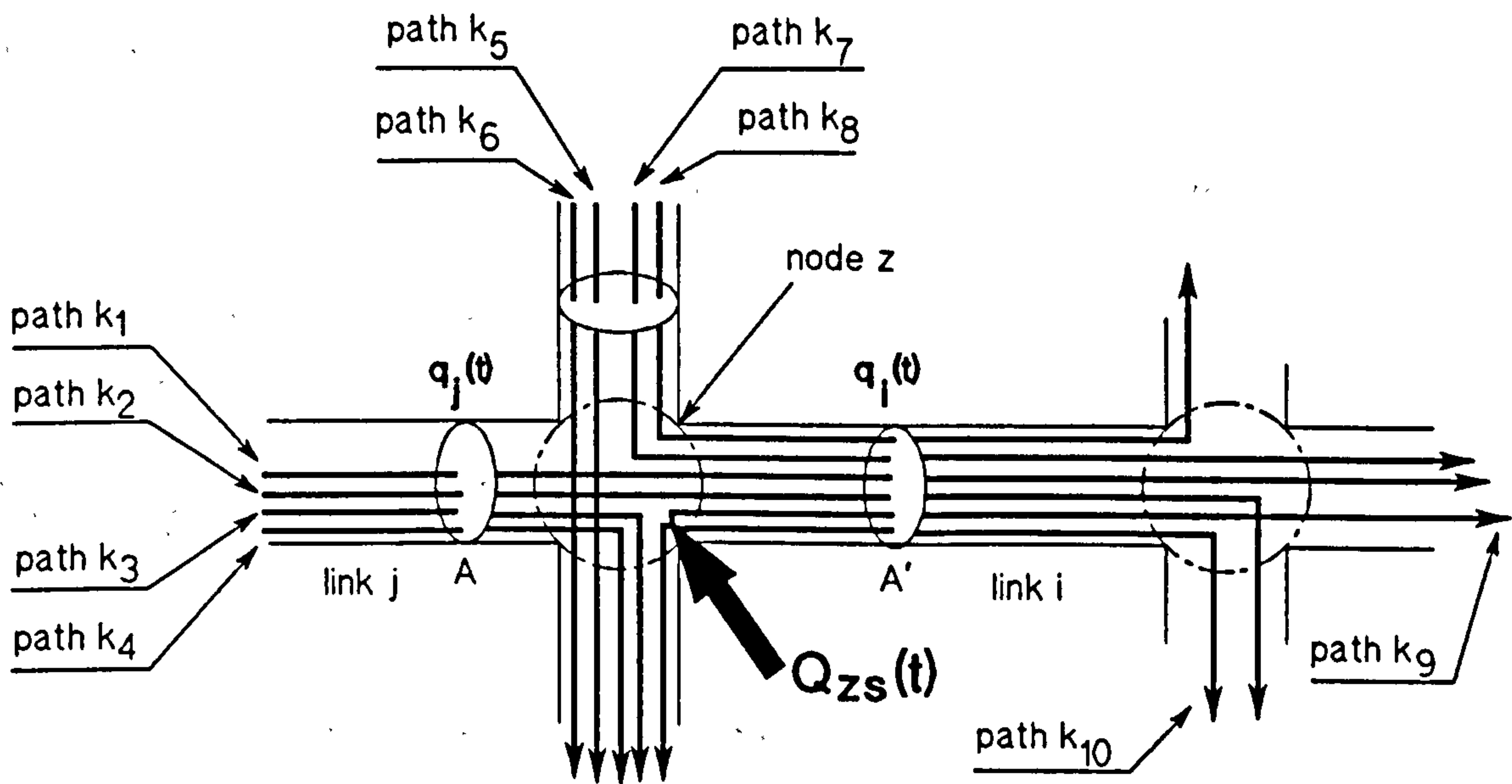


Figure 6.3 : Different components of flow within a link.

(ii) the speed of a vehicle within a link is assumed to be constant. Thus the time needed for a vehicle entering link  $i$  at time  $t$ , to traverse that link, denoted by  $tt_i(t)$  is expressed as :

$$tt_i(t) = d_i/v_i(t) \quad (6.6)$$

where

- $d_i$  is the length of link  $i$ , and
- $v_i(t)$  is the speed of the vehicle within link  $i$  at time  $t$ .

Since the traffic conditions within a link are assumed to be homogeneous, the flow  $q_i(t)$  and therefore the link outflow  $LOUT_i(t)$ , can be expressed using elementary relationships from traffic flow theory. Thus :

$$q_i(t) = LOUT_i(t) = K_i(t).v_i(t) \quad (6.7)$$

where  $K_i(t)$  is the concentration (density) within link  $i$  at time  $t$ , defined as :

$$K_i(t) = X_i(t)/d_i \quad (6.8)$$

where  $X_i(t)$  is the number of vehicles within link  $i$  at time  $t$ .

Substituting equation (6.8) in (6.7) the link flow (and outflow) can be expressed as :

$$q_i(t) = \text{LOUT}_i(t) = X_i(t) \cdot v_i(t) / d_i \quad (6.9)$$

The outflow,  $\text{LOUT}_j(t)$ , from all the links  $j$  that terminate at a node  $z$ , contribute to the inflow,  $\text{NIN}_z(t)$ , of that node at time  $t$ . If a node serves as an origin, the inflow to this node at time  $t$  is increased by the number of trips that depart from that origin node at time  $t$  :

$$\text{NIN}_z(t) = \sum_{j \in \Lambda(z)} \text{LOUT}_j(t) + \sum_{s \in S(z)} Q_{zs}(t) = \sum_{j \in \Lambda(z)} q_j(t) + \sum_{s \in S(z)} Q_{zs}(t) \quad (6.10)$$

where

$\Lambda(z)$  is the set of links that terminate at node  $z$ ,

$S(z)$  is the set of destination nodes such that  $\forall s \in S(z)$ ,  $z-s$  is an O-D pair.

In order the continuity of flow to be maintained, a *node flow conservation* condition must be satisfied. This requires that the inflow to each node at time  $t$  must be equal to the outflow from this node at the time  $t$ . i.e.

$$\text{NIN}_z(t) = \text{NOUT}_z(t) \quad (6.11)$$

In other words, since nodes are not associated with any impedance to flow (i.e. travel time), the number of vehicles entering a node  $z$  at time  $t$  must be equal to the number of cars leaving that node at time  $t$ . The outflow from a node contributes to the inflow of all the links emanating from this node. If a node also serves as a destination, a part of its outflow exits from the network. Furthermore, the outflow  $\text{LOUT}_j(t)$  can be also assumed to consist of the different flow components corresponding to the paths traversing this link, and the distribution of a node outflow to the links emanating from that node is performed assuming that the flow passing through a node remains in a disaggregate form. Thus, the component  $p_{j,k}(t) \cdot \text{LOUT}_j(t)$  of the outflow  $\text{LOUT}_j(t)$  leaving link  $j$  at time  $t$ , and following route  $k$ , will enter link  $i$  if  $i$  is the link following  $j$  along route  $k$ , in other words if  $i, j \in L_{k,rs}$  and  $O_k(i) = O_k(j) + 1$

Thus, let  $i$  be a link emanating from node  $z$ . Then the inflow to link  $i$  at time  $t$ , denoted by  $\text{LIN}_i(t)$ , consists of

- different outflow components coming from all the links terminating to node  $z$ , and corresponding to the paths which include link  $i$  at time  $t$ , and

- The components of the travel demand which enter the network from node z, and follow paths which include link i.

$$LIN_i(t) = \sum_{j \in L(i)} \sum_{rs} \sum_{k \in K_{rs}} LOUT_j(t) \cdot p_{j,k}(t) \cdot \delta_{i,k} \cdot \delta_{j,k} + \sum_{s \in S(z)} \sum_{k \in K_{zs}(t)} Q_{k,zs}(t) \cdot \phi_{i,k} \quad (6.12)$$

where

$L(i)$  is the set of links that terminate at the node from which link i emanates.

$\delta_{i,k}$  is an indicator variable such that i denotes a link and k denotes a path, and which is equal to 1 when link i is included in path k, and 0 otherwise.

$\phi_{i,k}$  is an indicator variable which is equal to 1 when link i is the first element in the ordered set  $L_{k,rs}$  (i.e., when  $O_k(i) = 1$ ), and 0 otherwise.

Thus in the example depicted in figure 6.3, the flow  $q_i(t)$ , passing a point A' in link i at time t, is the aggregation of the flows corresponding to paths  $k_1, k_2$ , (coming from link j),  $k_7, k_8$ , and  $k_9, k_{10}$  (originating from node z, i.e. due to  $Q_{k,zs}(t)$ ).

Therefore, because the OUT patterns of the links preceding a node contribute to the IN patterns of that node, and the OUT patterns of a node contribute to the IN patterns of succeeding links, traffic is moved through the network. Furthermore, the inflow to a link at time t represents the number of vehicles entering the link at time t, while the outflow the number of cars leaving that link. Consequently, a *link volume conservation* condition must be satisfied. This condition, expressed as :

$$\frac{dX_i(t)}{dt} = LIN_i(t) - LOUT_i(t) \quad (6.13)$$

implies that the rate of change of the number of vehicles within link i can be expressed as the difference per unit of time between the number of vehicles that enter the link (inflow) and the number of vehicles that leaving that link (outflow).

Using equations (6.12) and (6.13) the link volume conservation condition is given by the following equation :

$$\frac{dX_i(t)}{dt} = \sum_{j \in L(i)} \sum_{rs} \sum_{k \in K_{rs}} LOUT_j(t) \cdot p_{j,k}(t) \cdot \delta_{i,k} \cdot \delta_{j,k} + \sum_{rs} \sum_{k \in K_{rs}(t)} Q_{k,rs}(t) \cdot \phi_{i,k} - LOUT_i(t) \quad (6.14)$$

Substituting equation (6.9) into (6.14) the latter is expressed as:

$$\begin{aligned} \frac{dX_i(t)}{dt} = & \sum_{j \in L(i)} \sum_{rs} \sum_{k \in K_{rs}} \frac{v_j(t) \cdot X_j(t)}{d_j} \cdot p_{j,k}(t) \cdot \delta_{i,k} \cdot \delta_{j,k} + \\ & + \sum_{rs} \sum_{k \in K_{rs}(t)} Q_{k,rs}(t) \cdot \phi_{i,k} - \frac{v_i(t) \cdot X_i(t)}{d_i} \end{aligned} \quad (6.15)$$

### 6.3.2 The speed - density relationships

The link volume conservation condition (equation (6.15)) developed in the previous subsection provides the basic relationship which can be used to estimate the evolution over time of the traffic patterns. This relationship includes the variable  $v_i(t)$  representing the vehicles' velocity within a link which however depends on the link density, and therefore on the number of vehicles within the link. In this subsection three different speed - density models are presented. The first two were proposed by de Palma *et al.* (1984), while the third is developed from a travel time model which has been widely used in static assignment procedures.

The first relationship is based on a deterministic queueing model. A bottleneck is assumed to exist in each link  $i$ , so that the outflow from the link cannot exceed the capacity  $c_i$  of the bottleneck. Thus in the case that the bottleneck is not congested, the speed in link  $i$  is constant :

$$v_i(t) = v_i \quad (6.16)$$

where  $v_i$  is the constant speed in link  $i$  in the case that there is no congestion. In the case of congestion, however, the outflow of the link is constraint by the capacity  $c_i$ . Thus the outflow in this deterministic queue is given by :

$$LOUT_i(t) = q_i(t) = \min[ c_i, v_i X_i(t)/d_i ] \quad (6.17)$$

Using equation (6.9), and the above expression, the following model provides the speed in a link  $i$  at time  $t$  :

$$v_i(t) = \frac{\min[ c_i d_i, v_i X_i(t) ]}{X_i(t)} \quad (6.18)$$

The second model is based on elementary relationships between speed and concentration taken from traffic flow theory. The speed in a link is thus assumed to be a decreasing function of the traffic density which has the following form :

$$v_i(t) = v_i / \{ 1 + a[X_i(t)/D_i]^b \} \quad (6.19)$$

where

- $v_i$  is the free flow speed in link  $i$
- $D_i$  is the length of the road network within link  $i$ ,  
i.e.,  $D_i = (\text{no of lanes}) \times (\text{length of link} = d_i)$ . The values of  $D_i$  in this model play the same role as the values of the capacity in the first model.
- $a, b$  are nonnegative constants. The values for  $a$  and  $b$  considered by de Palma *et al.* (1984) were  $0.09 (\text{veh/km})^{-0.5}$  and  $0.5$  respectively.

As was discussed in chapter 3, the most widely used travel time model in static traffic assignment is the BPR volume delay curve (eq. 3.4) already presented in section 3.3. This model having the form

$$tt_i = tt_0 [ 1 + a(q_i/c_i)^b ] \quad (6.20)$$

can be used to derive a speed - volume relationship which can be used in DSUE network analysis. Thus introducing the relationship  $q_i(t) = X_i(t).v_i(t)/d_i$ , equation (6.20) can be transformed to :

$$\frac{d_i}{v_i(t)} = \frac{d_i}{v_i} \cdot \left[ 1 + a \cdot \left[ \frac{X_i(t).v_i(t)}{d_i.c_i} \right]^b \right] \quad (6.21)$$

and therefore  $v_i(t)$  can be defined by solving numerically the following equation :

$$v_i(t) = \frac{v_i}{\left[ 1 + a \cdot \left[ \frac{X_i(t).v_i(t)}{d_i.c_i} \right]^b \right]} \quad (6.22)$$



### 6.3.3 Formulation of the time dependent O-D travel times

The travel time associated with a route  $k$  which connects origin  $r$  to destination  $s$  is the sum of travel times on the links comprising this route. However since the link travel times are not constant during the period of analysis, the O-D travel time depends not only on the route a traveller has selected but also on the time he has departed. Thus, let  $TT_{k,rs}(t)$  denote the travel time from origin  $r$  to destination  $s$ , when departing at time  $t$  and using route  $k$ . This O-D travel time will be the sum of the time varying link travel times  $tt_i(t_i)$ , where  $i$  is a link along path  $k$ , and  $t_i$  denotes the time that a vehicle, following route  $k$  and departing at  $t$ , will traverse link  $i$ .

As was discussed earlier in this section the speed of a vehicle within a link is assumed to be constant. The constant speed of a vehicle entering link  $i$  at time  $t$  is taken as equal to  $v_i(t)$ , i.e., the speed calculated at the time that the vehicle enters the link.

The assumption regarding the speed of a vehicle (i.e., being constant) within a link seems to be unrealistic. However this assumption is not restrictive, since vehicles can be considered to move at variable speeds and  $v_i(t)$  to reflect the average speed experienced by the driver who enters link  $i$  at time  $t$ . Furthermore, it is reasonable to assume that the prevailing state within a link is 'first in - first out' or 'no overtaking' and therefore, the velocity of a vehicle is affected by the traffic density in front of it, i.e. the density which is created by earlier entries in the link.

Consequently, modelling the average vehicle speed,  $v_i(t)$ , as a function of the number  $X_i(t)$  of the vehicles within link  $i$  at time  $t$ , provides a realistic representation since the average speed of a vehicle within a link is assumed to be dependent on the number of cars which are distributed over the link at the time that the vehicle enters that link.

Having defined  $v_i(t)$  as the average vehicle speed, the travel time to traverse a link  $i$  for a vehicle which enters link  $i$  at time  $t$  is given by :

$$tt_i(t) = \frac{d_i}{v_i(t)} \quad (6.23)$$

Thus, since the link travel time depends on the time a vehicle enters the link, in order to calculate the O-D travel time  $TT_{k,rs}(t)$ , first the arrival time at each link along path  $k$  for a vehicle departing at time  $t$ , should be calculated.

Let  $AT_{i,k}(t)$  be the arrival (entry) time at link  $i$  for a vehicle departing at time  $t$  and following route  $k \in K_{rs}(t)$ . If  $i_1$  is the first link along route  $k$ , i.e., if  $i_1 \in L_{k,rs}$  and  $O_k(i_1)=1$ , then

$$AT_{i_1,k}(t) = t \quad t \in [T_0, T_0+T], i_1 \in L_{k,rs} \text{ and } O_k(i_1)=1 \quad (6.24)$$

The arrival time at the second link  $i_2$  will be given by :

$$AT_{i_2,k}(t) = t + tt_{i_1}(t) = AT_{i_1,k}(t) + tt_{i_1}(AT_{i_1,k}(t)) \quad (6.25)$$

where  $i_1, i_2 \in L_{k,rs}$ ,  $O_k(i_2) = O_k(i_1)+1$

Similarly the arrival at any link  $j$ , will be given by the following recursive formula :

$$AT_{j,k}(t) = AT_{i,k}(t) + tt_i(AT_{i,k}(t)) \quad i,j \in L_{k,rs} \text{ and } O_k(i) = O_k(j) - 1 \quad (6.26)$$

Since, the time needed to cross a node is assumed to be zero, the time that a driver, who departed at time  $t$  and used route  $k$ , will enter link  $j$  is equal to the time that he left from link  $i$  ; this time is denoted by  $LT_{i,k}(t)$ , and is expressed as :

$$LT_{i,k}(t) = AT_{i,k}(t) + tt_i(AT_{i,k}(t)) \quad (6.28)$$

The arrival time at the destination, for a driver who travels from  $r$  to  $s$ , follows route  $k$  and departs at time  $t$ , denoted by  $TD_{k,rs}(t)$ , can then be defined as :

$$TD_{k,rs}(t) = AT_{y,k}(t) + tt_y(AT_{y,k}(t)) \quad (6.29)$$

or equivalently as :

$$TD_{k,rs}(t) = LT_{x,k}(t) + tt_y(LT_{x,k}(t)) \quad (6.30)$$

where  $y$  is the link along route  $k \in K_{rs}(t)$  which is connected to the destination, i.e. ,  $y$  is the last element of the ordered set  $L_{k,rs}$  ,  $x \in L_{k,rs}$  and  $O_k(x) = O_k(y) - 1$

The O-D travel time from  $r$  to  $s$  for a driver departing at time  $t$  and following path  $k$ , can be computed as:

$$TT_{k,rs}(t) = TD_{k,rs}(t) - t \quad (6.31)$$

Having defined  $LT_{i,k}(t)$ , the variable  $p_{j,k}(t)$  used in the volume conservation equation (6.14) is calculated as follows :

$$p_{j,k}(t) = \frac{Q_{k,rs}(T^*_{j,k}(t))}{\sum_{rs} \sum_{m \in K_{rs}} Q_{m,rs}(T^*_{j,m}(t)) \cdot \delta_{j,m}} \quad (6.32)$$

where  $T^*_{j,k}(t)$  is the time that a vehicle, following path  $k$ , has to depart from the origin in order to leave link  $j$  at time  $t$ .  $T^*_{j,k}(t)$  can be defined as the following relationship exists :

$$LT_{j,k}(T^*_{j,k}(t)) = t \Rightarrow T^*_{j,k}(t) = LT_{j,k}^{-1}(t) \quad (6.33)$$

#### 6.3.4. The queueing theory approach

The most conventional queueing theory, namely the *steady state* queueing theory, assumes random vehicle arrivals and service patterns but no time variation in demand or capacity; furthermore it predicts infinite queue lengths whenever demand equals capacity. In reality, however, when demand approaches or exceeds capacity for short or moderate time periods, the growth of the queue length is less than the value predicted by the steady-state theory, because the queue takes time to grow; the practical result is that queue lengths always remain finite. The *deterministic time dependent queueing* approach takes these effects into account and provides a more realistic description of the growth and decay of queues in situations where the demand flow changes with time. This is particularly important during peak periods where the demand flow rises to a level which is close to (or even exceeds) capacity for a short time. However, deterministic theory assumes regular traffic arrivals and predicts zero queue length until demand exceeds capacity. In practice, because of the random nature of vehicular arrivals, there is always some probability of a queue even when demand is well below capacity. Thus, ideally, a time dependent stochastic queueing model for example like the one proposed by Catling (1977) would provide the most realistic representation of queue lengths and waiting times. However, this type of model would further complicate the problem and would substantially increase the computational costs.

Thus, because of the time varying nature of traffic flow it is necessary to use time dependent queueing theory to calculate queue lengths and vehicle delays. Furthermore, the *fluid approximation* approach, (developed by Newell, 1971), which will be adopted here, assumes that for large queues the discrete and stochastic arrival and departure process can be approximated by nonstochastic and continuous variables.

In the queueing model developed in this thesis, every link  $i$  in the network is assumed to have the structure depicted in figure 6.4, i.e., it consists of two different segments :

- $A_iC_i$  with an adequate capacity so that it is never congested and free flow conditions are always held
- $C_iD_i$  which is a bottleneck of fixed capacity  $c_i$

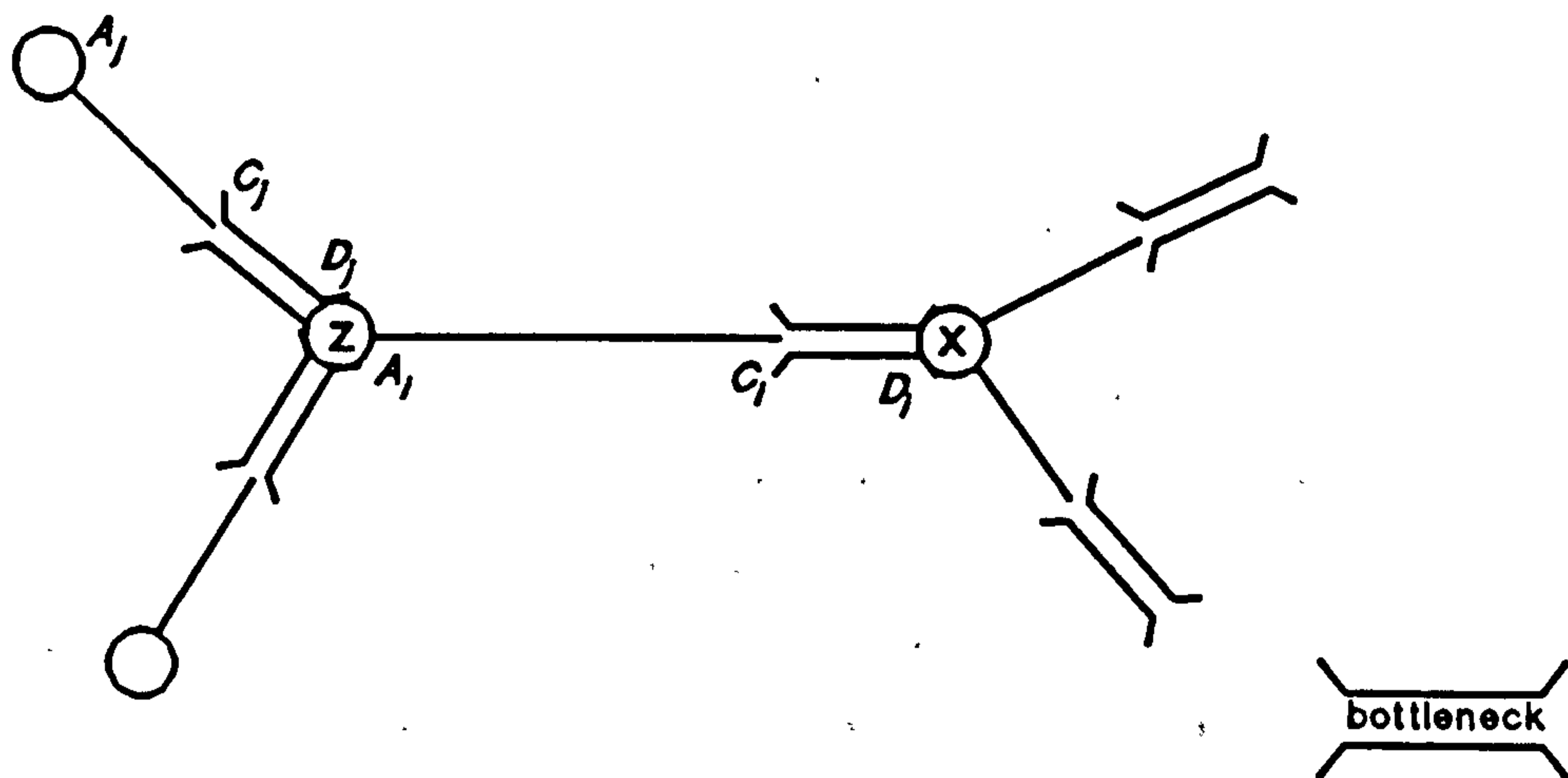


Figure 6.4 : Structure of network links used in the queueing theory approach.

Thus the travel time from  $A_i$  to  $C_i$  is constant and is denoted by  $to_i$ , while when the arrival rate at  $C_i$  exceeds the capacity  $c_i$ , a queue may develop.

The travel time from  $A_i$  to  $C_i$  is always constant whether there is a queue or not. In other words the queue length is assumed to be very short relative to the length of the segment  $A_iC_i$ , as if the vehicles were stacked up at point  $C_i$ . This assumption was used in all the existing DSUE models where queueing theory was applied in order to calculate link travel times.

Let  $tw_i(t)$  be the waiting time for a driver arriving at  $C_i$  at time  $t$ . Then for a driver entering link  $i$  at time  $t$ , the travel time needed to traverse link  $i$  is :

$$tt_i(t) = to_i + tw_i(t + to_i) \quad (6.34)$$

Let denote by  $Q_i(t)$  the queue length expressed as the number of cars in the queue formed at bottleneck  $C_i$  at time  $t$ , and let  $AR_i(t)$  be the arrival rate at  $C_i$  at time  $t$ . Then the rate of change in the number of cars waiting at the queue at time  $t$  is derived from a model of a deterministic queue as follows :

$$\frac{dQ_i(t)}{dt} = \begin{cases} AR_i(t) - q_i(t) & \text{for congestion} \\ 0 & \text{for no congestion} \end{cases} \quad (6.35)$$

where  $q_i(t)$  expresses the rate of outflow (departure rate) from link  $i$  at time  $t$ . Since the capacity of the bottleneck is considered fixed and a deterministic approach was adopted  $q_i(t)$  is defined as:

$$q_i(t) = \begin{cases} c_i & \text{for congestion} \\ AR_i(t) & \text{for no congestion} \end{cases} \quad (6.36)$$

Following the analysis of the traffic flow model presented in subsection 6.3.1, the arrival rate  $AR_i(t)$  is assumed to consist of an aggregation of different arrival rate components coming from the links,  $j$ , terminating at the node,  $z$ , from which link  $i$  emanates. Each of these components expresses a departure rate component of a link  $j \in \Lambda(z)$  and is specific to a route  $k$ , where  $k$  satisfies the condition  $\delta_{j,k} \cdot \delta_{i,k} = 1$ . Furthermore, since the travel time from  $A_j$  to  $C_i$  is assumed to be constant, an arrival at  $C_i$  at time  $t$  is equivalent to a departure from  $B_j$  at time  $t - t_{0j}$ .

The time dependent arrival rate can be therefore calculated from the following equation which has a form similar to equation (6.12) :

$$AR_i(t) = \sum_{j \in L(i)} \sum_{rs} \sum_{k \in K_{rs}} q_j(t-t_{0j}) \cdot p_{j,k}(t_j^*(t-t_{0j})) \cdot \delta_{i,k} \cdot \delta_{j,k} + \sum_{z \in S(z)} \sum_{k \in K_{zs}(t)} Q_{k,zs}(t-t_{0j}) \cdot \phi_{i,k} \quad (6.37)$$

where

$t_j^*(t)$  is the time that a vehicle departing from bottleneck  $j$  at time  $t$ , has joined the queue formed at  $C_j$ .  $t_j^*(t)$  can be defined by solving numerically the following relationship :

$$\begin{cases} t_j^*(t) + t w_j(t_j^*(t)) = t & \text{for congestion} \\ t_j^*(t) = t & \text{for no congestion} \end{cases} \quad (6.38)$$

$p_{j,k}(t)$  is the proportion of the vehicles arriving at the end of the queue formed at bottleneck  $C_j$  at time  $t$  which follow path  $k$ .  $p_{j,k}(t)$  is given by the following formula :

$$p_{j,k}(t) = \begin{cases} 0 & \text{for } \delta_{j,k} = 0 \\ \frac{Q_{k,rs}(t-t_{0j})}{AR_j(t)} & \text{for } \phi_{i,k} = 1 \\ \frac{q_m(t-t_{0j}) \cdot p_{m,k}(t_m^*(t-t_{0j}))}{AR_j(t)} & \text{for } \delta_{m,k} \cdot \delta_{j,k} = 1, O_k(j) > 1, \text{ and} \\ & O_k(m) = O_k(j) - 1 \end{cases} \quad (6.39)$$

Using equations (6.35), (6.36) and (6.37) the rate of change of the queue length can be expressed as :

$$\frac{d Q_i(t)}{d t} = \sum_{j \in L(i)} \sum_{rs} \sum_{k \in K_{rs}} q_j(t-t_{0j}) \cdot p_{j,k}(t_j^*(t-t_{0j})) \cdot \delta_{i,k} \cdot \delta_{j,k} + \sum_{rs} \sum_{k \in K_{rs}(t)} Q_{k,rs}(t-t_{0i}) \cdot \phi_{i,k} - q_i(t) \quad (6.40)$$

The waiting time for a driver arriving at  $C_i$  at time  $t$  is given by the following model of a deterministic queue:

$$tw_i(t) = \begin{cases} Q_i(t) / c_i & \text{for congestion} \\ 0 & \text{for no congestion} \end{cases} \quad (6.41)$$

Once the time varying link travel times (6.34) are calculated using the queueing model, the O-D travel times,  $TT_{k,rs}(t)$ , can be defined following the procedure presented in subsection 6.3.3.

#### 6.4 The demand model

Drivers travelling between a certain O-D pair  $r$ - $s$  are faced with a two-dimensional choice problem, i.e., what time  $t \in [T_0, T_0+T]$  to depart from their origin and which route  $k \in K_{rs}$  to select in order to reach their destination. Following the analysis of multiple dimension choice behaviour, presented in section 2.6, travellers are assumed to establish a hierarchy between the two different dimensions of choice, (time and route) and then follow a sequential decision making process in order to select the alternative which they believe will maximise their utility of travel.

Existing models analysing route and departure time choice in single O-D pair networks connected by parallel routes (Ben-Akiva et al. (1986a,b)), suggest the choice hierarchy depicted in figure 6.5 ; this choice hierarchy will be adopted in this thesis.

The two sets of alternatives available to an individual are :

$T = \{t_1, t_2, \dots, t_n\} = \{ \text{all possible departure times} \}$

$K_{rs} = \{k_1, k_2, \dots, k_m\} = \{ \text{all possible routes connecting } r\text{-}s \text{ for all departure times} \}$

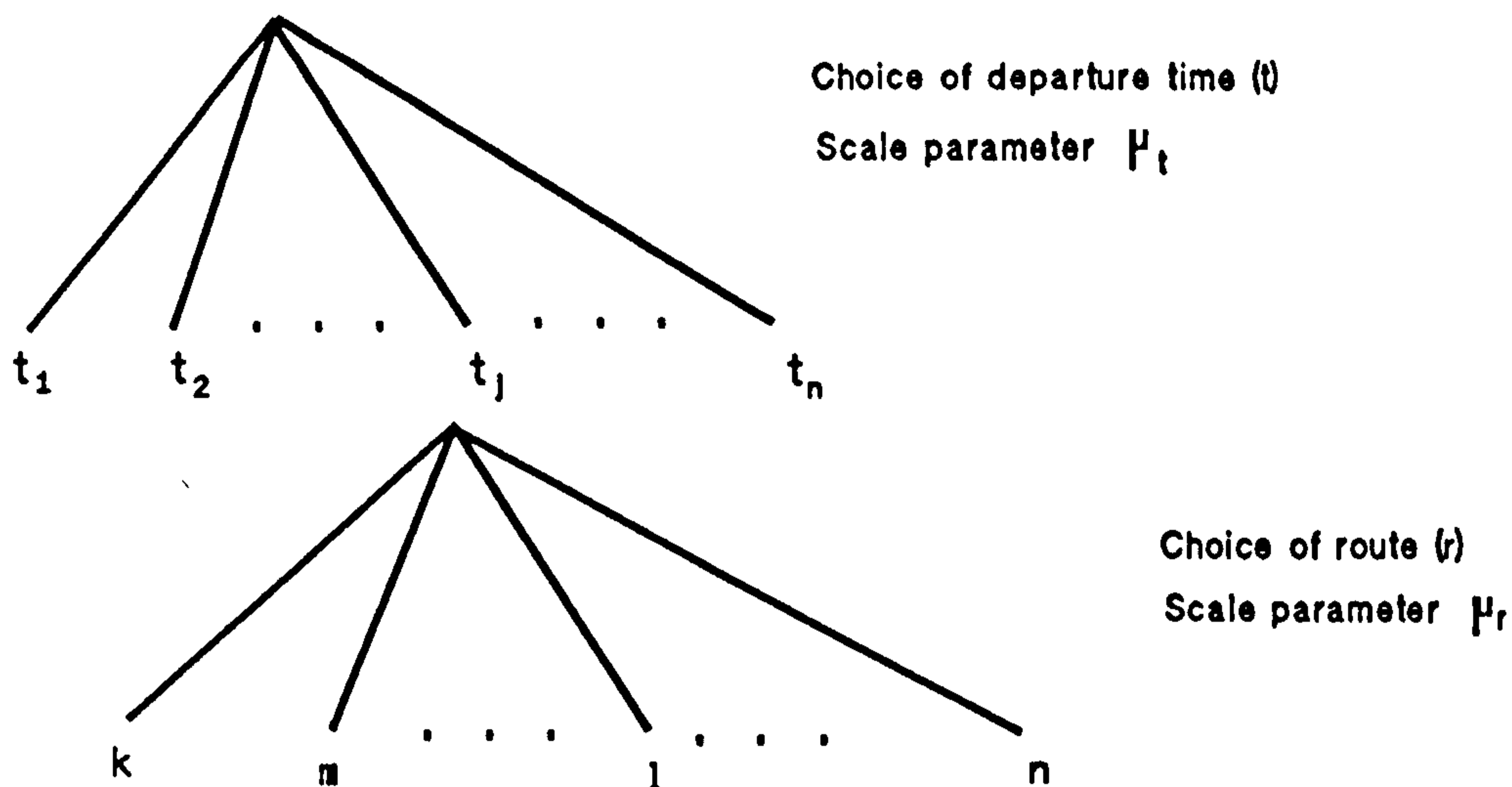


Figure 6.5: The choice hierarchy used in the development of the DSUE model.

Let  $T = \{t_1, t_2, \dots, t_n\}$  denote also the set of feasible departure times. The set of feasible routes however, as was argued in section 6.2, does not remain the same for all the alternative departure times; instead each departure time is associated with a set of feasible routes.

Thus, the choice set including all the feasible combinations of departure time and route will be defined as :

$$\bigcup_{t \in T} t \times K_{rs}(t) = \{(t_1, k) | k \in K_{rs}(t_1)\} \cup \{(t_2, k) | k \in K_{rs}(t_2)\} \cup \dots \dots \dots \cup \{(t_n, k) | k \in K_{rs}(t_n)\} \quad (6.42)$$

The adopted hierarchical structure of the alternatives implies that an individual will first decide on what time to depart and then subject to his choice he will select which route to follow. In other words, the probability that he will select route  $k$  and departure time  $t$ , denoted by  $P_{k,rs}(t)$ , is expressed as:

$$P_{k,rs}(t) = (\text{Probability of a departure at time } t) \cdot (\text{Probability of selecting route } k \text{ given a departure at time } t)$$

Travellers are assumed to make their decisions based on the utility maximisation principles. Furthermore each probability in the above expression is assumed to have the multinomial logit form with its own scale parameter, reflecting the degree of heterogeneity of preferences among individual travellers.

Thus the probability that a driver, travelling from  $r$  to  $s$ , will depart at time  $t$  and select route  $k$ , is obtained from the following mixed discrete/continuous nested logit model :

$$P_{k,rs}(t) = \frac{e^{\mu_r V_{k,rs}(t)}}{\sum_{k \in K_{rs}(t)} e^{\mu_r V_{k,rs}(t)}} \cdot \frac{e^{\mu_t V_{rs}^*(t)}}{\int_{T_0}^{T_0+T} e^{\mu_t V_{rs}^*(u)} du} \quad (6.43)$$

where

$V_{k,rs}(t)$  is the measured utility experienced by a driver travelling from  $r$  to  $s$ , who departs at time  $t$  and selects route  $k$ .

$\mu_t, \mu_r$  are the scale parameters associated with the upper level of decision, (i.e., departure time choice) and lower level of decision, (i.e., route choice) respectively.  $\mu_t, \mu_r \rightarrow \infty$  implies a deterministic choice, while  $\mu_t, \mu_r \rightarrow 0$  a pure random choice.

$V_{rs}^*(t)$  is a composite variable defined as:

$$V_{rs}^*(t) = \frac{1}{\mu_r} \ln \sum_{k \in K_{rs}(t)} e^{\mu_r V_{k,rs}(t)} \quad (6.44)$$

and expresses the expected maximum utility from the choice of among the alternative feasible routes at time  $t$ .

A necessary condition which must be satisfied in order the above formulation to be valid is provided by equation (2.48) :

$$\frac{\mu_t}{\mu_r} \leq 1 \quad (6.45)$$

for  $(\mu_t/\mu_r) = 1$  the nested logit formulation given by equation 6.43 reduces to a multinomial logit form and there is no need for the composite cost to be calculated.

The nested logit formulation given by equation (6.43) can be obtained following the analysis of multidimensional choice behaviour presented in section 2.6. Equation (6.43) can be obtained from equations (2.36), (2.45), (2.47) and (2.51), assuming that the utility function (eq. (2.37)), denoted by  $U_{k,rs}(t)$ , has the form :

$$U_{k,rs}(t) = V_{k,rs}(t) + e(t) + e_{k,rs}(t)$$

where  $e(t)$  is the random element attributable to departure time  $t$ , and  $e_{k,rs}(t)$  the random element attributable to the combination of route  $k$  and departure time  $t$ .



There are two possible interpretations of the error term that is incorporated in the utility function; these are clarified below:

In the nested logit formulation which is presented above and is used for predicting the traffic flow patterns in the analysis that will be presented in subsequent chapters, the utility associated with an alternative is considered as a random variable, in order to model the different utility levels that different travellers attribute to a certain alternative. The probabilistic distribution of the random utility, therefore represents the distribution of the actual utility experienced by different travellers, and the error term is attributed to the utility associated with unmeasured attributes of the alternatives, or in other words it reflects real utility which however cannot be measured by the analyst. On the other hand the mean utility ( $= V_{k,rs}(t)$  which is used in the nested logit formulation) represents the component of the utility which can be measured. Thus, in the case that the error term is used in order to represent the component of the real utility experienced by travellers, that cannot be measured, the higher the variance of the error term is, the wider the distribution of the actual utility experienced by the travellers will be, and therefore the higher the maximum expected utility will be.

In the case that the concept of utility is used to represent the perceived utility associated with a certain alternative, the mean utility  $V_{k,rs}(t)$  represents the actual utility associated with an alternative and which can be measured, whereas the error term reflects the travellers' perception errors. Under these assumptions, the composite cost derived from a set of reasonable alternatives is not expected to increase as the variance of the perception error increases, and can be expressed as the weighted average:

$$V_{rs}^*(t) = \sum_k S_{k,rs}(t) \cdot V_{k,rs}(t)$$

where  $S_{k,rs}(t)$  is the share of the trips between  $r$  and  $s$  at time  $t$  that are associated with route  $k$ .

A composite cost formulation which is based on weighted averages is used for evaluating the performance of the network that is carried out in chapter 9.

The demand model having the form of the mixed discrete/continuous nested logit expressed by equation (6.43) provides the probability of selecting a certain departure time and route, given the distribution of the measured utility  $V_{k,rs}(t)$  function. This function will be analysed in the following subsection.

### 6.4.1 Specification of the utility function

As was discussed in chapter 4, the main attributes of a trip that vary among alternative combinations of departure time and route, and which influence travellers' choices are (i) travel time and (ii) schedule delay. These are the main sources of disutility and therefore they must be included in the utility function  $V_{k,rs}(t)$ .

#### Disutility of travel time

The disutility of travel increases as the travel time increases. Nonlinear functional forms might provide the most realistic representation of the travel time disutility. However for simplicity, in this work the linear form depicted in fig. 6.6, which was used in existing DSUE models (section 5.3.2), will be adopted.

The marginal disutility of an additional unit of travel time is  $\alpha > 0$ . Thus the component of the total utility of travel which is associated with travel time is expressed as:

$$-\alpha \cdot TT(t) \quad (6.46)$$

#### Disutility of schedule delay.

Let  $[t_{rs} - D_{rs}, t_{rs} + D_{rs}]$ , where  $D_{rs} > 0$ , be the desired time period of arrival at the destination  $s$ , for a driver travelling between  $r$ - $s$ . This period expresses a type of indifference band (used in the bounded rationality user equilibrium).  $t_{rs}$  denotes the centre of that period and  $D_{rs}$  reflects a measure of work start time flexibility. Assuming that  $TD_{k,rs}(t)$  denotes the time that a driver, who departed at time  $t$  and used route  $k$ , will arrive at his destination  $s$ , then :

$$TD_{k,rs}(t) = t + TT_{k,rs}(t) \quad (6.47)$$

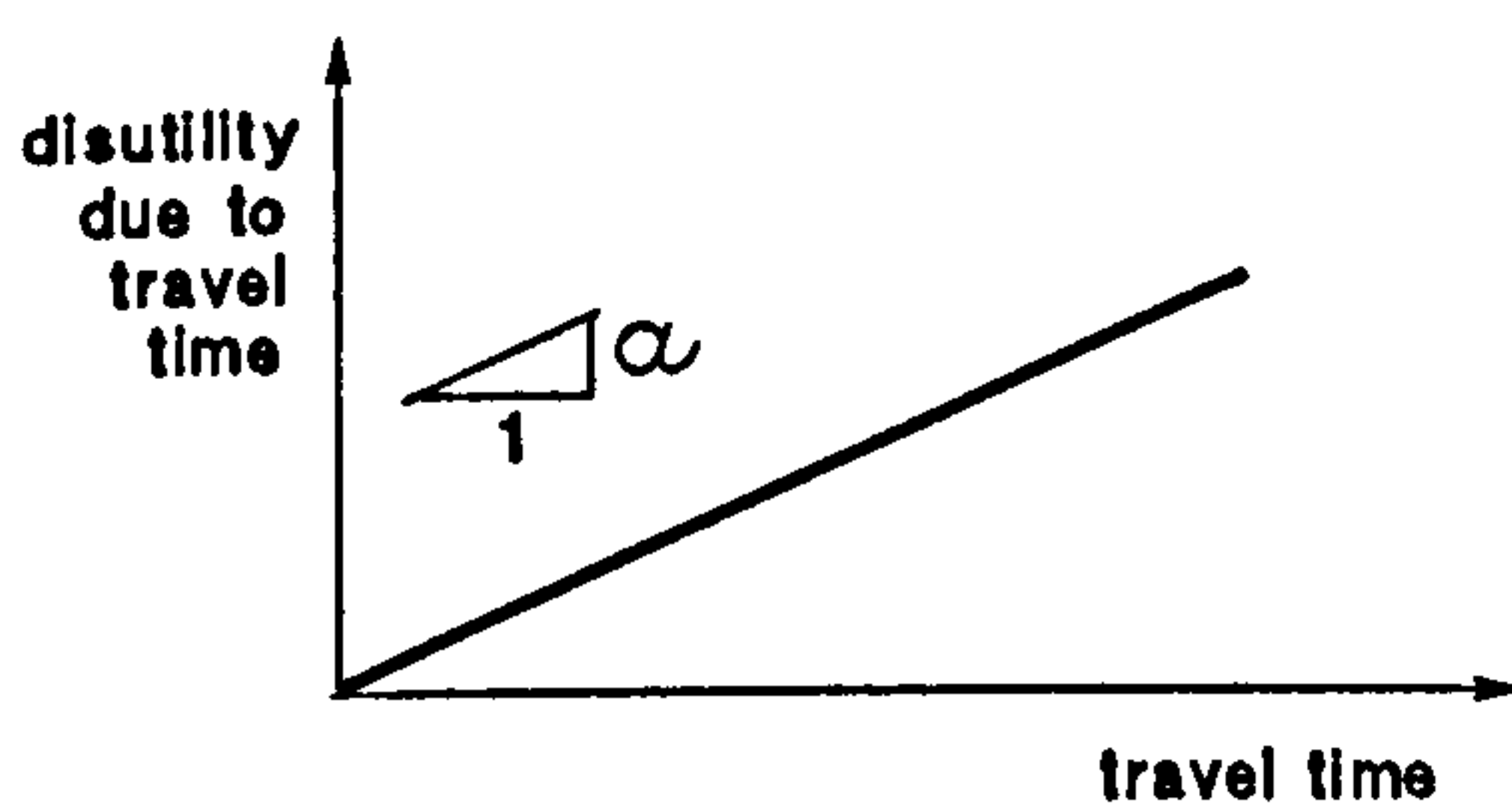


Figure 6.6a: Disutility due to travel time

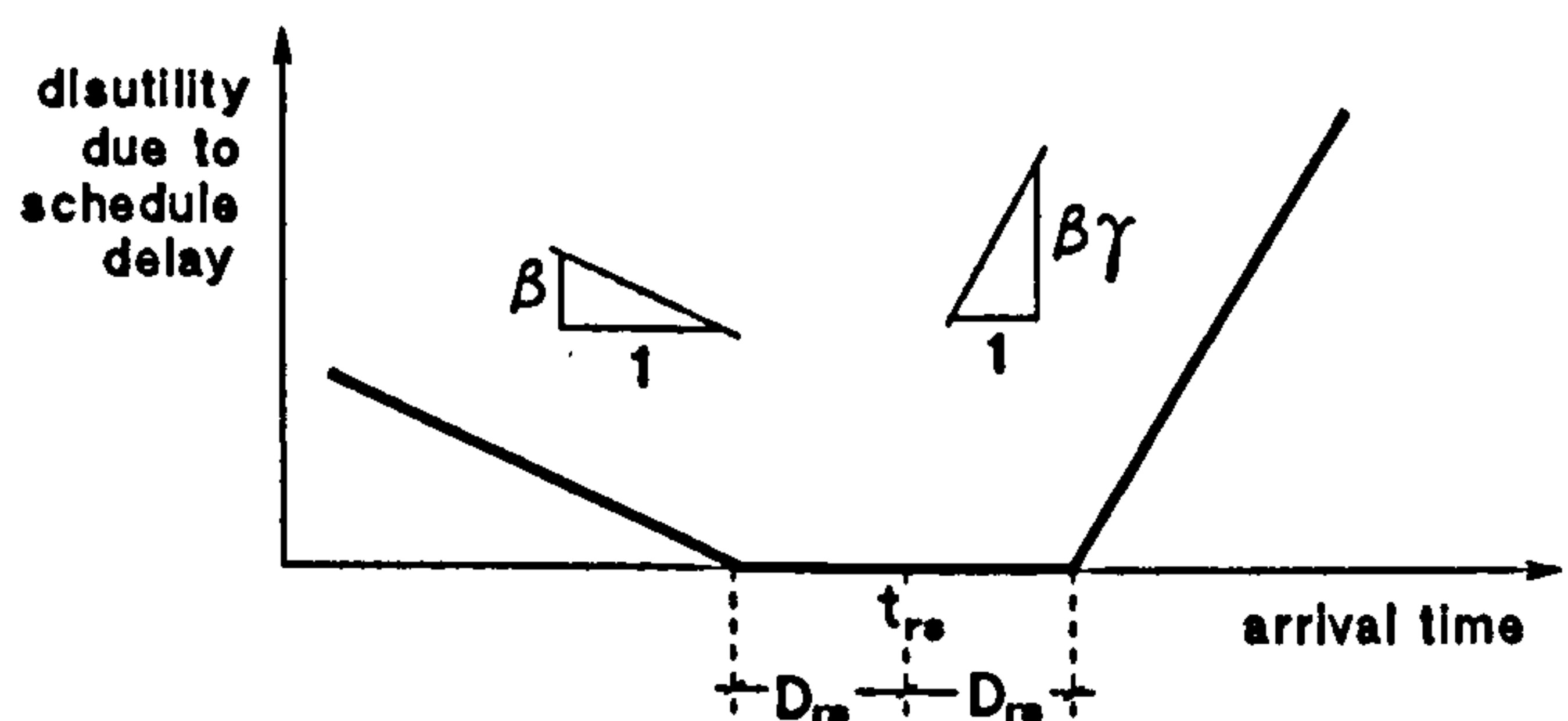


Figure 6.6b: Disutility due to schedule delay vs arrival time at destination

Let  $t_{k,rs}^1$  and  $t_{k,rs}^2$  be the departure times from the origin  $r$ , associated with route  $k$  and O-D pair  $r$ - $s$ , such that :

$$TD_{k,rs}(t_{k,rs}^1) = t_{rs} - D_{rs} \quad (6.48)$$

$$TD_{k,rs}(t_{k,rs}^2) = t_{rs} + D_{rs} \quad (6.49)$$

Thus, for the drivers travelling between the O-D pair  $r$ - $s$  and following route  $k$ , departures from  $r$

- before the time point  $t_{k,rs}^1$  are early departures,
- during the interval  $(t_{k,rs}^1, t_{k,rs}^2)$  are on-time departure, and
- after the time point  $t_{k,rs}^2$  are late departures.

Using equations (6.47), (6.48) and (6.49),  $t_{k,rs}^1$  and  $t_{k,rs}^2$  are defined as :

$$t_{k,rs}^1 = t_{rs} - D_{rs} - TT_{k,rs}(t_{k,rs}^1) \quad (6.50)$$

$$t_{k,rs}^2 = t_{rs} + D_{rs} - TT_{k,rs}(t_{k,rs}^2) \quad (6.51)$$

In other words,  $t_{k,rs}^1$  denotes the earliest possible departure time which a traveller following route  $k$  can select, in order to arrive on time at his destination. Similarly,  $t_{k,rs}^2$  denotes the latest possible departure time that will enable a driver following route  $k$  to arrive at his destination within his desired period of arrival.

Having defined  $t_{k,rs}^1$  and  $t_{k,rs}^2$ , another variable which will be useful in the formulation of the utility function can be introduced. This is denoted by  $\theta_{k,rs}(t)$  and is expressed as:

$$\theta_{k,rs}(t) = \begin{cases} 1 & \text{for early arrivals, i.e., for } t \leq t_{k,rs}^1 \\ 0 & \text{for on time arrivals, i.e., for } t_{k,rs}^1 < t < t_{k,rs}^2 \\ -\gamma & \text{for late arrivals, i.e., for } t \geq t_{k,rs}^2 \end{cases} \quad (6.52)$$

The disutility due to schedule delay depends on the arrival time at destination and the desired period of arrival. A graphical representation of the relationship between the disutility of schedule delay and arrival time, which was already used (section 5.3.2) and which will be adopted in this work, is depicted in figure 6.6b . For an arrival earlier than  $t_{rs} - D_{rs}$  the disutility of schedule delay is assumed to be a linearly decreasing function of the arrival time; the rate of decrease is denoted by  $\beta > 0$ . For an arrival later than  $t_{rs} + D_{rs}$ , the disutility is assumed to increase linearly at a rate  $\beta.\gamma$ , where  $\gamma > 0$ . The utility associated with schedule delay can then be expressed as:

$$\left\{ \begin{array}{ll} -\beta \cdot [(t_{rs} - D_{rs}) - (t + TT_{k,rs}(t))] & \text{for } \theta_{k,rs}(t) = 1 \\ 0 & \text{for } \theta_{k,rs}(t) = 0 \\ -\beta \cdot \gamma \cdot [(t + TT_{k,rs}(t) - (t_{rs} + D_{rs}))] & \text{for } \theta_{k,rs}(t) = -\gamma \end{array} \right. \quad (6.53)$$

and using the variable  $\theta_{k,rs}(t)$ , this utility can take the form :

$$\beta \cdot D_{rs} \cdot |\theta_{k,rs}(t)| - \beta \cdot \theta_{k,rs}(t) \cdot [t_{rs} - t - TT_{k,rs}(t)] \quad (6.54)$$

### Total utility of travel

The total of travel includes the utility due to travel time and the utility due to schedule delay and it is given by the sum of equations (6.45) and (6.53):

$$V_{k,rs}(t) = -\alpha \cdot TT_{k,rs}(t) + \beta \cdot D_{rs} \cdot |\theta_{k,rs}(t)| - \beta \cdot \theta_{k,rs}(t) \cdot [t_{rs} - t - TT_{k,rs}(t)] \quad (6.55)$$

### 6.4.2 The time dependent departure rates

Once the utility function is defined, the probability that an individual, travelling from  $r$  to  $s$ , will select route  $k$  and departure time  $t$ , can be calculated. The departure rate  $Q_{k,rs}(t)$  is then defined as:

$$Q_{k,rs}(t) = Q_{rs} \cdot P_{k,rs}(t) = Q_{rs} \cdot \frac{e^{\mu_r V_{k,rs}(t)}}{\sum_{k \in K_{rs}(t)} e^{\mu_r V_{k,rs}(t)}} \cdot \frac{e^{\mu_t V_{rs}^*(t)}}{\int_{T_0}^{T_0+T} e^{\mu_t V_{rs}^*(u)} du} \quad (6.56)$$

where  $Q_{rs}$  denotes the total number of trips between the O-D pair  $r$ - $s$ , and  $P_{k,rs}(t)$  is given by equation (6.43).

In the above equation the departure rate  $Q_{k,rs}(t)$  is expressed as a function of the distribution of the utility function  $V_{k,rs}(t)$ ,  $\forall k,rs,t$ , which in turn is directly related to the distribution of the O-D travel times  $TT_{k,rs}(t)$  (eq. (6.55)). The time dependent O-D travel time can be calculated from the link travel times,  $tt_i(t)$ , following the procedure described in subsection 6.3.3. Link travel times however are dependent on link flows (for the traffic flow model) or queue lengths, which can be defined only if the departure rate distributions  $Q_{k,rs}(t)$  are known (eq. (6.15) and (6.40)). The right hand side of equation is therefore a function of the variables  $Q_{k,rs}(t)$ ,  $\forall k,rs,t$ . This is because the proposed DSUE model consists of (i) a travel time model which calculates the O-D travel times  $TT_{k,rs}(t)$  given the departure rate distributions  $Q_{k,rs}(t)$

and (ii) a demand model which calculates the variables  $Q_{k,rs}(t)$ ,  $\forall k,rs,t$ , given the  $TT_{k,rs}(t) \forall k,rs,t$ .

The complexity of the relationships involved in the formulation of both the demand and the travel time model do not allow the derivation of analytical solutions. However, the model as specified mainly by equations (6.26), (6.29), (6.30), (6.55) and (6.56), and (6.34), (6.40) and (6.41) for the queueing model, or (6.15) and (6.18 or 6.19 or 6.22) for the traffic flow model, can be solved iteratively by using a dynamic framework which describes the evolution of the departure rate distributions over time. This procedure, termed the demand adjustment mechanism, is developed in the following section.

### 6.5 The demand adjustment mechanism

In this section a dynamic extension of the demand model is presented, based on the dynamic logit formulation which was discussed in chapter 2.

It is assumed that decisions are made at the individual level on the basis of the utility maximisation principle; travellers are willing to minimise their perceived disutility of travel, which however depends not only on their own decisions, but also on the decisions made by the whole population of travellers who use the transportation system. Thus, individuals do not behave independently on each other, but there is an interaction between users' behaviour which can be attributed to the fact that the system's performance is demand dependent. Choice decisions are therefore not static, but evolve over time.

The dynamic formulation represents the interaction between individuals' decisions as it is directed by their own criteria of choice and the transportation system's characteristics. It describes the evolution of the time-of-day dependent departure rate and travel time distributions over time. Thus, it can also predict the transient travel patterns after the implementation of a new policy to relief congestion or during temporary traffic restrictions.

A static choice model assumes the existence of an equilibrium state that can be considered as the stationary solution of a dynamic process. Thus the steady state solution of the dynamic model, which is described in this section, will provide the dynamic stochastic user equilibrium state.

The setting of the system is the same as before, with additional notation to indicate the day-to-day variability. Thus all the variables used in the travel time model and the

demand model, are functions of  $t$ , the departure time from the origin, as well as  $\omega$ , the day for which they defined, e.g.  $V_{k,rs}(t,\omega)$ ,  $tt_i(t,\omega)$ ,  $TT_{k,rs}(t,\omega)$ ,  $Q_{k,rs}(t,\omega)$ .

Users of a transportation network continuously modify their trip decisions based on the information they acquire from recent trips. It is assumed that the next time a tripmaker will use the transportation system, he will either change his current trip choices and search for a better option, or he will remain at his current decision state.

Let us define by  $F(m,t' | V(t,\omega), \omega)\Delta\omega$  the probability that during the time interval  $(\omega, \omega+\Delta\omega)$ , an individual who departed during  $(t', t'+\delta t')$  and selected route  $m$ , decides to review his current trip choices, given the global distribution of the utility,  $V(t, \omega)$ . This distribution is however defined from the geometric characteristics of the system, and the global trip choice distribution of the total population of travellers,  $Q(t)$ . It is therefore the evolution, over time, of the time dependent departure rate distributions which requires a detailed analysis.

Thus consider the O-D pair  $r$ - $s$ , and let denote by  $F_{m,k}(t',t,\omega)\Delta\omega\delta t$  the fraction of individuals who shift from a departure during  $[t', t'+\delta t]$  to a departure during  $[t, t+\delta t]$ , and switch from route  $m \in K_{rs}(t', \omega)$  to route  $k \in K_{rs}(t, \omega)$  during the time interval  $[\omega, \omega+\Delta\omega]$ . The rate of change of the number of individuals following route  $k$  and departing during the interval  $[t, t+dt]$ , can then be expressed as the difference per unit of time between the number of individuals shifting to  $[t, t+dt]$  and switching to  $k$ , and the number of individuals shifting from  $[t, t+dt]$  and/or switching from  $k$ . This rate of change can be therefore expressed as follows :

$$\begin{aligned} \frac{\partial Q_{k,rs}(t,\omega)}{\partial \omega} &= \sum_{t'} \sum_{m} Q_{m,rs}(t',\omega) \cdot F_{m,k}(t',t,\omega) - \\ &\quad (t',m) \neq (t,k) \\ &\quad - Q_{k,rs}(t,\omega) \sum_{t'} \sum_{m} F_{k,m}(t,t',\omega) \end{aligned} \quad (6.57)$$

Let us denote by  $F_{k|c}(t,\omega)$  the probability that a traveller will select route  $k$  and departure time  $t$ , given that he decided to change his current trip decisions. Then the transition rate from the decision state  $[t',m]$  to the state  $[t,k]$  can be expressed as :

$$F_{m,k}(t',t,\omega) = F(m,t' | V(t,\omega), \omega) \cdot F_{k|c}(t,\omega) \quad (6.58)$$

Before presenting the formulae which calculate the conditional probabilities  $F_{k|c}(t,\omega)$ , a simplifying assumptions used in the analysis of the transition rates between different decision states is introduced. Thus following Ben-Akiva *et al.* (1984) the utility of a

shift to a new decision state is assumed to be independent of the attributes of the current state implying that there is a constant transition rate out of the current decision state. Thus the probability that an individual, who departed during  $(t', t'+dt')$  and selected route  $m$  at time  $\omega$ , decides to review his current trip choices, is a constant,  $F_1$ , i.e. :

$$F(m,t' | V(t,\omega), \omega) = F_1 \quad \forall m, t', \omega \quad (6.59)$$

In other words  $F_1$  defines the probability that a randomly selected traveller will readjust his current trip choices.

Figure 6.7 illustrates the current and the possible next choices of the travellers, who will change their current trip decisions. These travellers can be classified in two categories :

(i) An individual,  $I_1$ , who belongs to the first category will alter his current trip decisions in both or at least in one of his dimensions of choice. Thus he may :

- switch to a different route  $k$ , and shift to a another departure time,  $t$ , (i.e., he will select the decision state  $[t,k]$ ), or

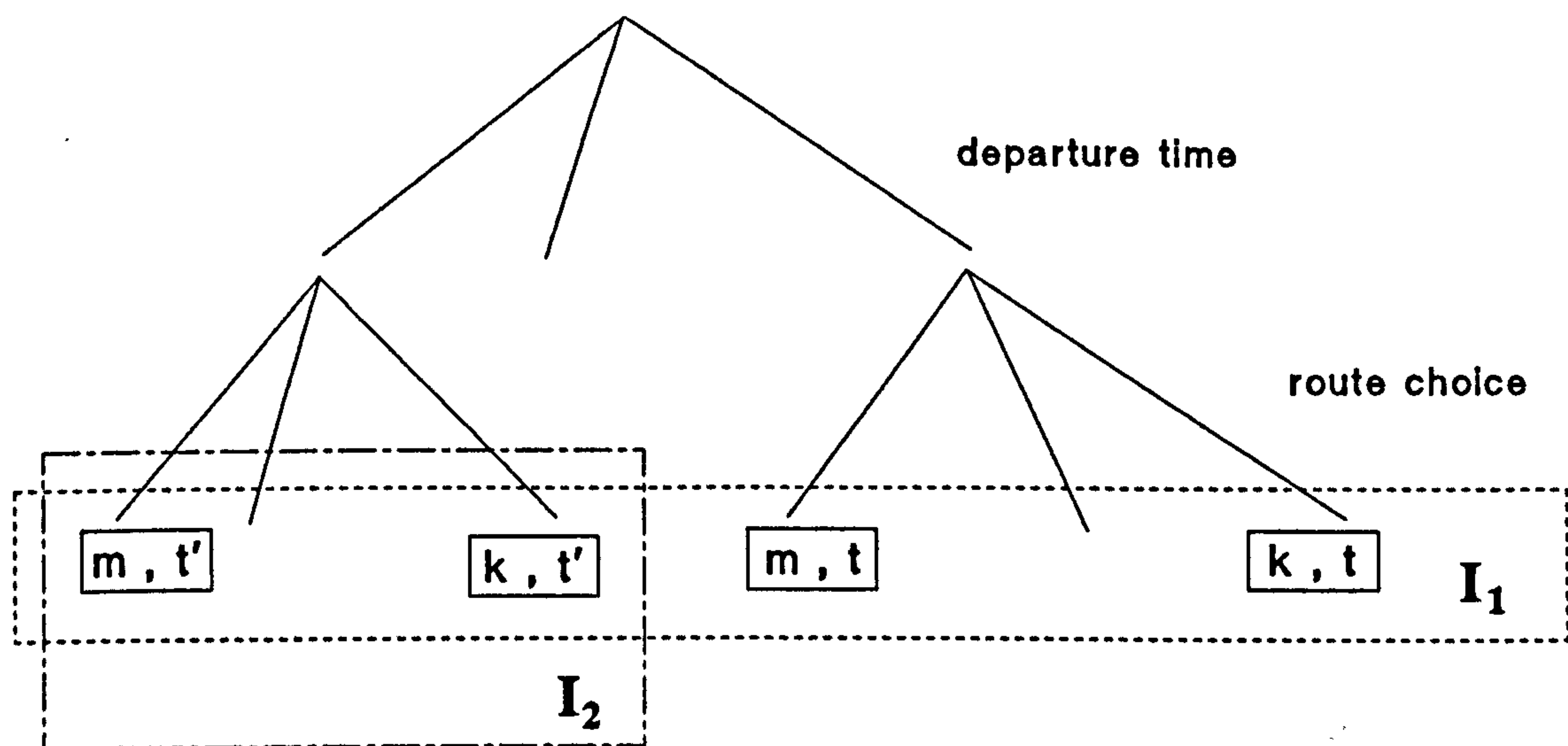


Fig. 6.7 : The different groups of individuals who review their trip choices

- switch to a different route  $k$ , and keep the same departure time  $t'$ , (i.e., he will shift to decision state  $[t',k]$ ), or
- shift to another departure time  $t$  and keep the same route,  $m$  (i.e., his new decision state will be  $[t,m]$ ).

(ii) A traveller  $I_2$  from the second category, will keep the same departure time,  $t'$ , and will switch to another route,  $k$ . In other words he will adopt a decision state  $[t',k]$ .

It is assumed that the ratio of the individuals  $I_1$  who belong to the first category, to the individuals  $I_2$  who belong to the second is constant. Thus the probability that a randomly selected traveller, who reviews his trip choices, will alter his current trip decisions in both or at least in one of his dimensions of choice is assumed to be constant  $F_2$ ; therefore the probability that he will keep the same departure time and switch to another route is equal to  $(1-F_2)$

Consider an individual travelling between the O-D pair  $r$ - $s$ . Once he decides to review his current decision state  $[t,k]$ , he compares the alternatives from his own choice set and is assumed to decide on the basis of the utility maximisation principle as that used in the logit formulation. Thus he first estimates the utility  $U_{k,rs}(t, \omega+\Delta\omega)$  associated with each alternative  $[t,k]$  at time  $\omega+\Delta\omega$ , and then chooses the alternative that maximises his utility.

A traveller  $I_1$  who belongs to the first category, is faced with a two dimensional choice problem : the choice of departure time and route which, he believes, will minimise his disutility of travel. The alternatives are assumed to be hierarchically structured, and the sequential decision making process implied is modelled using a nested logit. The probability that a randomly selected traveller from the first category, will select a route  $k$  and a departure time  $t$ , is given by the following equation which has the form of the nested logit given by equation (6.43)

$$FF_{k|c1}(t, \omega) = \frac{e^{\mu_r V_{k,rs}(t, \omega+\Delta\omega)}}{\sum_{k \in K_{rs}(t, \omega+\Delta\omega)} e^{\mu_r V_{m,rs}(t, \omega+\Delta\omega)}} \cdot \frac{e^{\mu_t V_{rs}^*(t, \omega+\Delta\omega)}}{\int_{T_0}^{T_0+T} e^{\mu_t V_{rs}^*(u, \omega+\Delta\omega)} du} \quad (6.60)$$

An individual  $I_2$  who belongs to the second category is faced with a single dimension choice decision problem, since he is assumed to keep the same departure time, and try to maximise his utility by readjusting his route choice only. Thus the probability that a randomly selected traveller from the second category, will select a route  $k$ , is expressed with the following multinomial logit :



$$FF_{k|c_2}(t', \omega) = \frac{e^{\mu_r V_{k,rs}(t', \omega + \Delta\omega)}}{\sum_{k \in K_{rs}(t, \omega + \Delta\omega)} e^{\mu_r V_{m,rs}(t, \omega + \Delta\omega)}} \quad (6.61)$$

The structure of the dynamic logit formulation, however, requires that travellers have perfect information regarding the levels of utility associated with different combinations of route and departure time, at time  $\omega + \Delta\omega$ . However, an individual can not obtain any such information, and he is therefore left with only one alternative; this is to estimate the temporal distribution of travel time and consequently utility levels at time  $\omega + \Delta\omega$ . It is suggested that each traveller assumes that the other commuters do not change their trip decision at time  $\omega + \Delta\omega$ , and thus the travel time distributions will remain the same as at  $\omega$ . The variable utility,  $V_{k,rs}(t, \omega + \Delta\omega)$ ,  $\forall t, k, rs$ , in equations (6.60) and (6.61) can be therefore expressed as a function of  $\omega$ , i.e., the latest time a traveller has used the system, instead of  $\omega + \Delta\omega$ , and defines the average utility experienced by the users who selected route  $k$  and departure time  $t$ , at time  $\omega$ .

On the other hand, complete information concerning travel times on alternative routes and for different departure times at time  $\omega$ , is also difficult to obtain. However users can increase their information from the knowledge acquired through experience with different routes and departure times, and by getting information from other drivers and media reports on the levels of congestion. Therefore, in the proposed demand adjustment mechanism, travellers are assumed to be perfectly informed, or to estimate the road conditions, during the latest period they used the transportation system, very accurately. Furthermore, the logit formulation adopted, is based on the random utility concept and therefore treats the utility as random variable, in order to take into account the perception errors involved at a decision making process.

As mentioned earlier in this chapter, due to the spatial and time variability of traffic patterns during the peak period, the set of reasonable paths for each O-D pair is not static but depends on the departure time from the origin. Furthermore, a path is considered to be reasonable, subject to the time dependent level of congestion (along the links it includes), which in turn is related to the temporal distribution of departure rates. However, trip decisions and thus departure rate distributions are not static but evolve over time. Therefore, the set of reasonable paths associated with a certain departure time will not remain the same over time but will be changing depending on the evolution of the departure rate distributions. This variability of the set of reasonable paths, gives rise to the definition of two new sets of routes (which will be

used later in this section) for each departure time  $t$ . The first, includes the routes which constitute a reasonable choice for a departure during the interval  $[t, t+\delta t]$  both at time  $\omega$  and at  $\omega+\Delta\omega$ ; it is denoted by  $DK_{rs}(t,\omega)$ , can be defined as :

$$DK_{rs}(t,\omega) = K_{rs}(t,\omega) \cap K_{rs}(t,\omega+\Delta\omega) \quad (6.62)$$

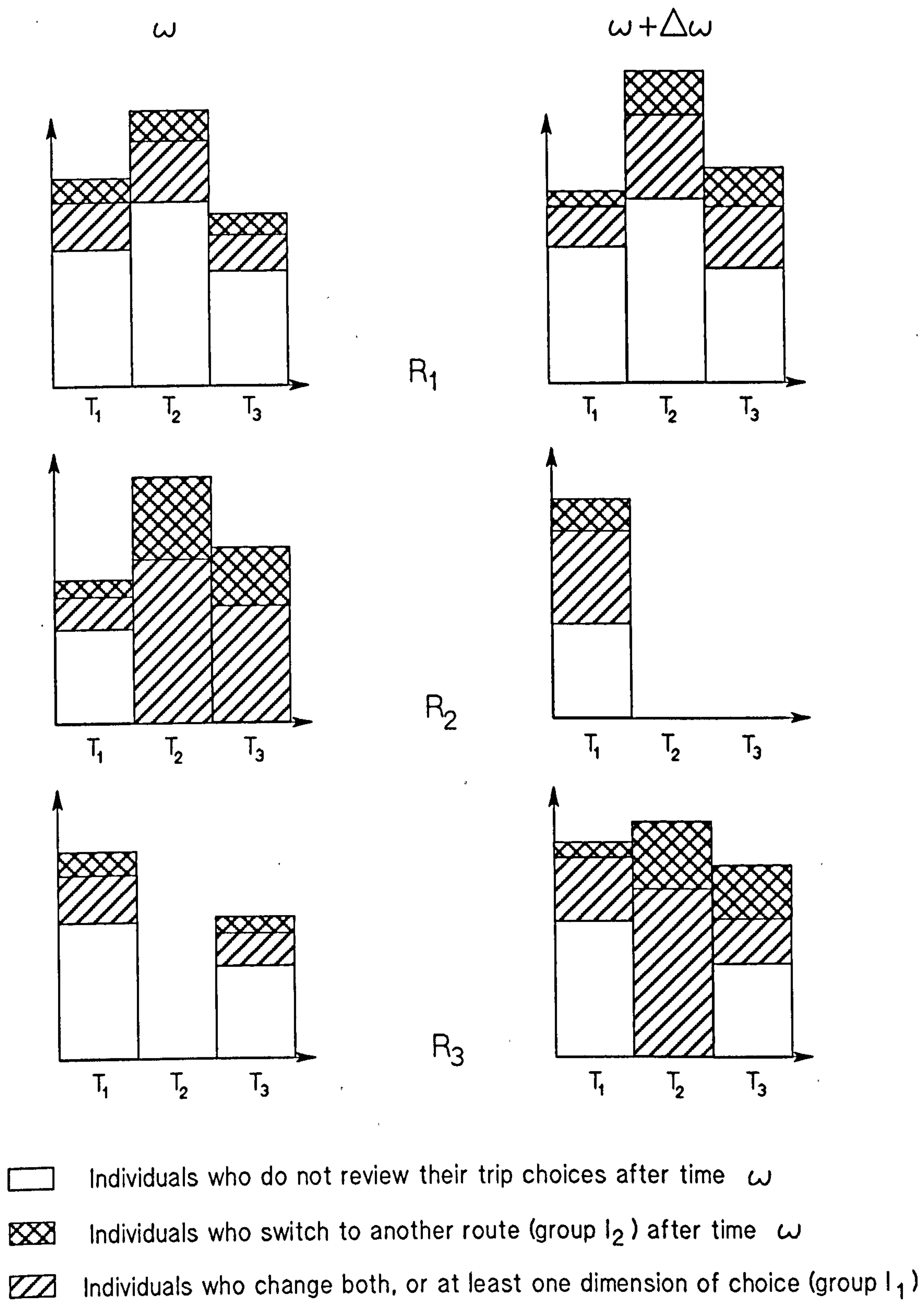
The second set includes the routes which are considered as reasonable choices for a departure during the interval  $[t, t+\delta t]$  at time  $\omega$  but not at  $\omega+\Delta\omega$ ; it is denoted by  $KK_{rs}(t,\omega)$ , can be defined as :

$$KK_{rs}(t,\omega) = K_{rs}(t,\omega) \sim K_{rs}(t,\omega+\Delta\omega) = \{ k \mid k \in K_{rs}(t,\omega) \wedge k \notin K_{rs}(t,\omega+\Delta\omega) \} \quad (6.63)$$

An individual first estimates the traffic patterns that will take place at time  $\omega+\Delta\omega$ , (which he assumes to be the same to the actual patterns occurred at time  $\omega$ ), and then defines the set  $K_{rs}(t,\omega+\Delta\omega)$ . It should be noted that  $K_{rs}(t,\omega+\Delta\omega)$  refers to the set of routes which are estimated as reasonable choices at time  $\omega+\Delta\omega$ , and not the ones which are actually defined as reasonable after the loading of the network at time  $\omega+\Delta\omega$ . This variable defines the choice set considered by a traveller at time  $\omega+\Delta\omega$ , and will therefore remain as a function of  $\omega+\Delta\omega$  in equations (6.60) and (6.61).

As stated earlier in this section, the probability that a randomly selected individual will review his current trip choices is assumed to be constant and equal to  $F_1$ , (eq. (6.59)). The value  $F_1$  is the same for all the travellers who have selected any combination of route  $k$  and departure time  $t$ , at time  $\omega$ , which remains a reasonable choice at time  $\omega+\Delta\omega$ , i.e. if  $k \in DK_{rs}(t,\omega)$ . Any individual who has selected a route  $k$  and a departure time  $t$ , such that  $k \notin DK_{rs}(t,\omega)$  will definitely have to review his current trip choices. Thus for the group of travellers who have selected route  $k \in KK_{rs}(t,\omega)$  the value of  $F_1$  will be equal to one.

To demonstrate how travellers readjust their trip decisions, consider the example illustrated in fig 6.8. The O-D pair examined is connected by three different routes,  $R_1$ ,  $R_2$ ,  $R_3$ , and for simplification reasons the period of analysis is assumed to consist of three successive time intervals  $T_1$ ,  $T_2$ ,  $T_3$ , during which demand is assumed to be constant. At time  $\omega$ , travellers estimate that routes  $R_1$  and  $R_2$  are reasonable choices during the whole peak period, while route  $R_3$  is not considered as a reasonable choice during the interval  $T_2$ . The path choice sets for each time interval, at time  $\omega$ , are therefore :



**Figure 6.8: Adjustments of trip choices made by travellers  $I_1$  and travellers  $I_2$ .**

$$K_{rs}(T_1, \omega) = \{ R_1, R_2, R_3 \}$$

$$K_{rs}(T_2, \omega) = \{ R_1, R_2 \}$$

$$K_{rs}(T_3, \omega) = \{ R_1, R_2, R_3 \}$$

The temporal distribution of departure rates at time  $\omega + \Delta\omega$ , has however led to an overloading of route 2, resulting in high travel times such that this route does not any more constitute a reasonable alternative during the intervals  $T_2$  and  $T_3$ .

On the other hand the overload of path 2 has resulted in a relief of the excess flow in some of the links included in path 3, which is now considered as a reasonable option during the interval  $T_2$ .

The path choice sets for each time interval, at time  $\omega + \Delta\omega$ , are therefore :

$$\left. \begin{array}{l} K_{rs}(T_1, \omega + \Delta\omega) = \{ R_1, R_2, R_3 \} \\ K_{rs}(T_2, \omega + \Delta\omega) = \{ R_1, R_3 \} \\ K_{rs}(T_3, \omega + \Delta\omega) = \{ R_1, R_3 \} \end{array} \right\} \Rightarrow \left\{ \begin{array}{ll} DK_{rs}(T_1, \omega) = \{ R_1, R_2, R_3 \}, & KK_{rs}(T_1, \omega) = \emptyset \\ DK_{rs}(T_2, \omega) = \{ R_1 \}, & KK_{rs}(T_1, \omega) = \{ R_2 \} \\ DK_{rs}(T_3, \omega) = \{ R_1, R_3 \}, & KK_{rs}(T_1, \omega) = \{ R_2 \} \end{array} \right.$$

Thus at time  $\omega + \Delta\omega$ , a fraction of  $F_1 \cdot F_2$  of the travellers who, at time  $\omega$ , have selected route  $R_1$  and  $R_3$  independent on departure time, and route  $R_2$  and departure time during the interval  $T_1$  will have to readjust their trip choices and select any alternative from the set  $\{ (R_1, T_1), (R_1, T_2), (R_1, T_3), (R_2, T_1), (R_3, T_1), (R_3, T_2), (R_3, T_3) \}$ . The same choice set will be considered by a fraction  $F_2$  of the travellers who used route  $R_2$  and departed during the intervals  $T_2$  and  $T_3$  (for the latter group of travellers  $F_1=1$ )

A fraction  $(1-F_2)$  of the individuals who have selected route  $R_2$  during the intervals  $T_2$  and  $T_3$  (for them  $F_1=1$ ), will keep the same departure time and will switch to either route  $R_1$  or  $R_3$ . Similarly, a fraction  $F_1 \cdot (1-F_2)$  of the individuals who have adopted a decision state  $[R_i, T_j]$  at time  $\omega$ , such that  $R_i \in DK_{rs}(T_j, \omega)$ ,  $j=1,2,3$ , (i.e., who selected a alternative which remains a reasonable choice at time  $\omega + \Delta\omega$ ) will depart at the same time,  $T_j$ , as at time  $\omega$  and select any route  $R_i \in K_{rs}(T_j, \omega + \Delta\omega)$

Having analysed the mechanisms directing the readjustment of trip choice, and using equations (6.58), (6.60) and (6.61), the transition rates from one decision state to another, can be expressed with the following set of equations.

$$F_{m,k}(t',t,\omega) = \left\{ \begin{array}{l} F_1 \cdot F_2 \cdot \frac{e^{\mu_r V_{k,rs}(t,\omega)}}{\sum_{m \in K_{rs}(t,\omega+\Delta\omega)} e^{\mu_r V_{m,rs}(t,\omega)}} \cdot \frac{e^{\mu_t V_{rs}^*(t,\omega)}}{\int_{T_0}^{T_0+T} e^{\mu_t V_{rs}^*(u,\omega)} du} + \\ + F_1 \cdot (1-F_2) \cdot \frac{e^{\mu_r V_{k,rs}(t,\omega)}}{\sum_{m \in K_{rs}(t,\omega+\Delta\omega)} e^{\mu_r V_{m,rs}(t,\omega)}} \quad \text{for } m \in DK_{rs}(t,\omega) \\ \\ F_2 \cdot \frac{e^{\mu_r V_{k,rs}(t,\omega)}}{\sum_{m \in K_{rs}(t,\omega+\Delta\omega)} e^{\mu_r V_{m,rs}(t,\omega)}} \cdot \frac{e^{\mu_t V_{rs}^*(t,\omega)}}{\int_{T_0}^{T_0+T} e^{\mu_t V_{rs}^*(u,\omega)} du} + \\ + (1-F_2) \cdot \frac{e^{\mu_r V_{k,rs}(t,\omega)}}{\sum_{m \in K_{rs}(t,\omega+\Delta\omega)} e^{\mu_r V_{m,rs}(t,\omega)}} \quad \text{for } m \in KK_{rs}(t,\omega) \end{array} \right. \quad (6.64)$$

Thus, using equations (6.57) and (6.64), and considering time as a continuous variable, the rate of change of the number of individuals following route  $k$  and departing during the interval  $[t, t+\delta t]$ , can be expressed as :

$$\begin{aligned}
 \frac{\partial Q_{k,rs}(t,\omega)}{\partial \omega} = & F_1 \cdot \left[ \int_{u=T_0}^{T_0+T} \sum_{m \in DK_{rs}(u,\omega)} Q_{m,rs}(u,\omega) \cdot F_2 \cdot \frac{e^{\mu_r V_{k,rs}(t,\omega)}}{\sum_{m \in K_{rs}(t,\omega+\Delta\omega)} e^{\mu_r V_{m,rs}(t,\omega)}} \cdot \frac{e^{\mu_t V_{rs}^*(t,\omega)}}{\int_{T_0}^{T_0+T} e^{\mu_t V_{rs}^*(w,\omega)} dw} du + \right. \\
 & \left. + \sum_{m \in DK_{rs}(t,\omega)} Q_{m,rs}(t,\omega) \cdot (1-F_2) \cdot \frac{e^{\mu_r V_{k,rs}(t,\omega)}}{\sum_{m \in K_{rs}(t,\omega+\Delta\omega)} e^{\mu_r V_{m,rs}(t,\omega)}} - Q_{k,rs}(t,\omega) \right] + \\
 & + \int_{u=T_0}^{T_0+T} \sum_{m \in KK_{rs}(u,\omega)} Q_{m,rs}(u,\omega) \cdot F_2 \cdot \frac{e^{\mu_r V_{k,rs}(t,\omega)}}{\sum_{m \in K_{rs}(t,\omega+\Delta\omega)} e^{\mu_r V_{m,rs}(t,\omega)}} \cdot \frac{e^{\mu_t V_{rs}^*(t,\omega)}}{\int_{T_0}^{T_0+T} e^{\mu_t V_{rs}^*(w,\omega)} dw} du +
 \end{aligned}$$

$$+ \sum_{m \in KK_{rs}(t, \omega)} Q_{m,rs}(t, \omega) \cdot (1 - F_2) \cdot \frac{e^{\mu_r V_{k,rs}(t, \omega)}}{\sum_{m \in K_{rs}(t, \omega + \Delta \omega)} e^{\mu_r V_{m,rs}(t, \omega)}} \quad (6.65)$$

The above equation defines the rate of change of the number of travellers departing during  $[t, t + \delta t]$  and following route  $k$ , for the case that  $k \in K_{rs}(t, \omega + \Delta \omega)$ . In the case that  $k \in KK_{rs}(t, \omega + \Delta \omega)$ , i.e., if  $k$  is not included in the set  $K_{rs}(t, \omega + \Delta \omega)$ , but included in the set  $K_{rs}(t, \omega)$ , then all the travellers who have adopted the decision state  $[t, k]$  at time  $\omega$  will alter their trip decisions, and no traveller will select the alternative  $[t, k]$  at time  $\omega + \Delta \omega$ . Therefore :

$$\frac{\partial Q_{k,rs}(t, \omega)}{\partial \omega} = - Q_{k,rs}(t, \omega) \quad \text{for } k \in KK_{rs}(t, \omega) \quad (6.66)$$

Equations (6.65) and (6.66) represent a set of nonlinear differential equations which describe the evolution of the departure rate distributions over time. However, since link flows, queue lengths and travel times can be defined given the departure rate distributions, the set of equations (6.65) and (6.66) determine the evolution over time of the whole system including flows and queues.

For  $\omega \rightarrow \infty$  this dynamic system is assumed to reach a stationary state where

$$\lim_{\omega \rightarrow \infty} V_{k,rs}(t, \omega) = V_{k,rs}(t) \quad \forall k, rs, t \quad (6.67)$$

$$\lim_{\omega \rightarrow \infty} Q_{k,rs}(t, \omega) = Q_{k,rs}(t) \quad \forall k, rs, t \quad (6.68)$$

where  $Q_{k,rs}(t)$  is given by equation (6.56).

However the complex structure of equation (6.65) does not allow the use of standard mathematical techniques to determine the uniqueness and stability properties of the system. Thus existence and uniqueness of the stability state of this model cannot be proved. However a large number of simulation experiments presented in chapter 8 have shown that in general, different initial conditions lead to the same steady state.

The demand adjustment mechanism presented in this section can be used with the demand and the travel time model presented in previous sections of this chapter to calculate the evolution of the traffic patterns over time. The model as presented so far,

has a theoretical form, which is rather not easy to apply. Based on the theoretical development, the next section derives a version of the model which is suitable for numerical simulation and derives the general framework of the simulation model.

### 3.7 The Framework of the Simulation Model

The dynamic simulation is essentially the numerical solution of the set of equations (6.65), and the set of equations comprising the travel time model presented in section 6.3.

Equation (6.65) describes the evolution over time of the departure rate distributions  $Q_{k,rs}(t,\omega)$  where time, denoted by  $\omega$ , is considered as a continuous variable. In the simulation model  $\omega$  is transformed into a discrete variable representing an iteration which corresponds to a day. Thus the discrete version of equation (6.65) describes the time evolution of the departure rate distributions from day  $\omega$  to day  $\omega+1$ .

Furthermore in the demand model trip choices are described by two variables (i) discrete route selection and (ii) continuous departure time. In the simulation framework of the model, the departure time  $t$  is also transformed into a discrete variable. Thus the time period  $[T_0, T_0+T]$  is divided into equal time intervals of length  $h$ , so that to represent the time unit for the discrete variable  $t$ . The variable  $Q_{k,rs}(t)$  is then used in order to express the number of travellers selecting route  $k$  and departing during the interval  $[t, t+h]$ .

Thus after transforming  $t$  and  $\omega$  into discrete variables, equation (6.65) is given by the following equation where the integrals over alternative departure times have been replaced by summations.

$$\begin{aligned}
 Q_{k,rs}(t,\omega+1) = & Q_{k,rs}(t,\omega) + \\
 & + F_1 \cdot \left[ \sum_{t'=T_0}^{T_0+T} \sum_{m \in DK_{rs}(t',\omega)} Q_{m,rs}(t',\omega) \cdot F_2 \cdot \frac{e^{\mu_r V_{k,rs}(t,\omega)}}{\sum_{m \in K_{rs}(t,\omega+1)} e^{\mu_r V_{m,rs}(t,\omega)}} \cdot \frac{e^{\mu_t V_{rs}^*(t,\omega)}}{\sum_{u=T_0}^{T_0+T} e^{\mu_t V_{rs}^*(u,\omega)}} + \right. \\
 & \left. + \sum_{m \in DK_{rs}(t,\omega)} Q_{m,rs}(t,\omega) \cdot (1-F_2) \cdot \frac{e^{\mu_r V_{k,rs}(t,\omega)}}{\sum_{m \in K_{rs}(t,\omega+1)} e^{\mu_r V_{m,rs}(t,\omega)}} - Q_{k,rs}(t,\omega) \right] +
 \end{aligned}$$

$$\begin{aligned}
& + \sum_{t'=T_0}^{T_0+T} \sum_{m \in K_{rs}} Q_{m,rs}(t',\omega) \cdot F_2 \cdot \frac{e^{\mu_r V_{k,rs}(t,\omega)}}{\sum_{m \in K_{rs}(t,\omega+1)} e^{\mu_r V_{m,rs}(t,\omega)}} \cdot \frac{e^{\mu_t V_{rs}^*(t,\omega)}}{\sum_{u=T_0}^{T_0+T} e^{\mu_t V_{rs}^*(u,\omega)}} + \\
& + \sum_{m \in K_{rs}(t,\omega)} Q_{m,rs}(t,\omega) \cdot F_1 \cdot (1-F_2) \cdot \frac{e^{\mu_r V_{k,rs}(t,\omega)}}{\sum_{u=T_0}^{T_0+T} e^{\mu_r V_{m,rs}(u,\omega)}} \quad (6.69)
\end{aligned}$$

The above equation defines the number of travellers using route  $k$  and departing at time  $t$  on day  $\omega+1$ , from the number of travellers with the same trip choices on day  $\omega$  and the time dependent utility levels  $V_{k,rs}(t,\omega)$  and departure rates  $Q_{k,rs}(t,\omega)$ , associated with each route  $k$  and departure time  $t$ .

During any interval  $[t, t+h]$  the traffic conditions are assumed to be constant, thus for example,  $v_i(t,\omega)$ ,  $X_i(t,\omega)$ ,  $Q_i(t,\omega)$  etc. remain constant during the interval  $[t, t+h]$ .

The discrete version of equation (6.15) can then be expressed as :

$$\begin{aligned}
X_i(t+h,\omega) = X_i(t,\omega) + h \cdot \left[ \sum_{j \in L(i)} \sum_{rs} \sum_{k \in K_{rs}} \frac{v_j(t,\omega) \cdot X_j(t,\omega)}{d_j} \cdot p_{j,k}(t,\omega) \cdot \delta_{i,k} \cdot \delta_{j,k} - \right. \\
\left. - \frac{v_i(t,\omega) \cdot X_i(t,\omega)}{d_i} \right] + \sum_{rs} \sum_{k \in K_{rs}(t)} Q_{k,rs}(t,\omega) \cdot \phi_{i,k} \quad (6.70)
\end{aligned}$$

The above expression implies that the number of cars,  $X_i(t+h,\omega)$ , within link  $i$  at time  $t+h$ , ( which is assumed to remain constant during the interval  $[t+h, t+2h]$ ) will be equal to the number of cars within link  $i$  at time  $t$ , increased by the number of cars entering  $i$  during the interval  $[t, t+h]$ , and decreased by the number of cars leaving link  $i$  during the same interval.

Similarly for the queueing model, the queue length can be expressed with the discrete version of (6.40) as follows :

$$\begin{aligned}
Q_i(t+h,\omega) = Q_i(t,\omega) + h \cdot \left[ \sum_{j \in L(i)} \sum_{rs} \sum_{k \in K_{rs}} q_j(t-t_{0i},\omega) \cdot p_{j,k}(t_j^*(t-t_{0i}),\omega) \cdot \delta_{i,k} \cdot \delta_{j,k} - q_i(t,\omega) \right] \\
+ \sum_{rs} \sum_{k \in K_{rs}(t)} Q_{k,rs}(t-t_{0i},\omega) \cdot \phi_{i,k} \quad (6.71)
\end{aligned}$$



The discrete version of the model as described above, can be solved iteratively when the initial conditions are known. The general structure of the simulation algorithm is shown in figure 6.9. The algorithm is an iterative process with an iteration corresponding to a day. The simulation starts with the specification of the parameters of the model and the initial conditions of the simulation include the time dependent set of reasonable paths  $K_{rs}(t,0)$  for each departure time  $t$ , and the distributions of the departure rates,  $Q_{k,rs}(t,0)$ , for each route and departure time, and all the O-D pairs, on day  $\omega = 0$ . As suggested by Ben-Akiva *et al.* (1986a,b), two cases can be considered for the initial departure rate distributions,  $Q_{k,rs}(t,0)$  :

- (i) They might be predetermined distributions, (e.g. uniform) or observed empirically.
- (ii) They are determined from the model as the steady state distributions under no congestion. This distribution labelled as the *pseudo-stationary state distribution* is the one derived from equation (6.56) assuming that travel times are equal to the free flow travel times.

The initial conditions for every day  $\omega$  consist of the link volumes  $X_i(T_0,\omega)$ , or the queue lengths  $Q_i(T_0,\omega)$ , at time  $T_0$ , for all the links of the network. Within each iteration, the departure rate distributions for day  $\omega+1$ , are computed from those of day  $\omega$ , using the utility levels on day  $\omega$ . Thus individuals are assumed to use information on traffic conditions during day  $\omega$  in order to make their choices on day  $\omega+1$ .

In every iteration, first the time dependent link volumes  $X_i(t,\omega)$ , or queue lengths  $Q_i(t,\omega)$  are calculated using equation (6.70) and (6.71). Link travel times are then defined from equations (6.18 or 19 or 22) and (6.23), or, (6.34) and (6.41). The link travel times are then used in order to define the set of reasonable paths  $K_{rs}(t,\omega+1)$ , following the procedure described in section 6.2. O-D travel times are defined using the equations in subsection 6.2.3 and the utility function is calculated for each alternative departure time and route from equation (6.55). The iteration ends with the estimation of the time dependent departure rates  $Q_{k,rs}(t,\omega+1)$ , for day  $\omega+1$ , by using equation (6.69).

The iterative process is terminated when either the last specified day of the simulation experiment is reached or when convergence is achieved. Convergence of the algorithm, or in other words an equilibrium solution is found when travellers do not any more alter their trip choices. The convergence criterion can be therefore expressed as:

$$RR = \underset{\substack{rs, k \in K_{rs} \\ t = T_0, \dots, T_0+T}}{\text{Max}} \frac{1}{F_1} \cdot \frac{|Q_{k,rs}(t,\omega+1) - Q_{k,rs}(t,\omega)|}{Q_{k,rs}(t,\omega)} \leq \xi \quad (6.72)$$

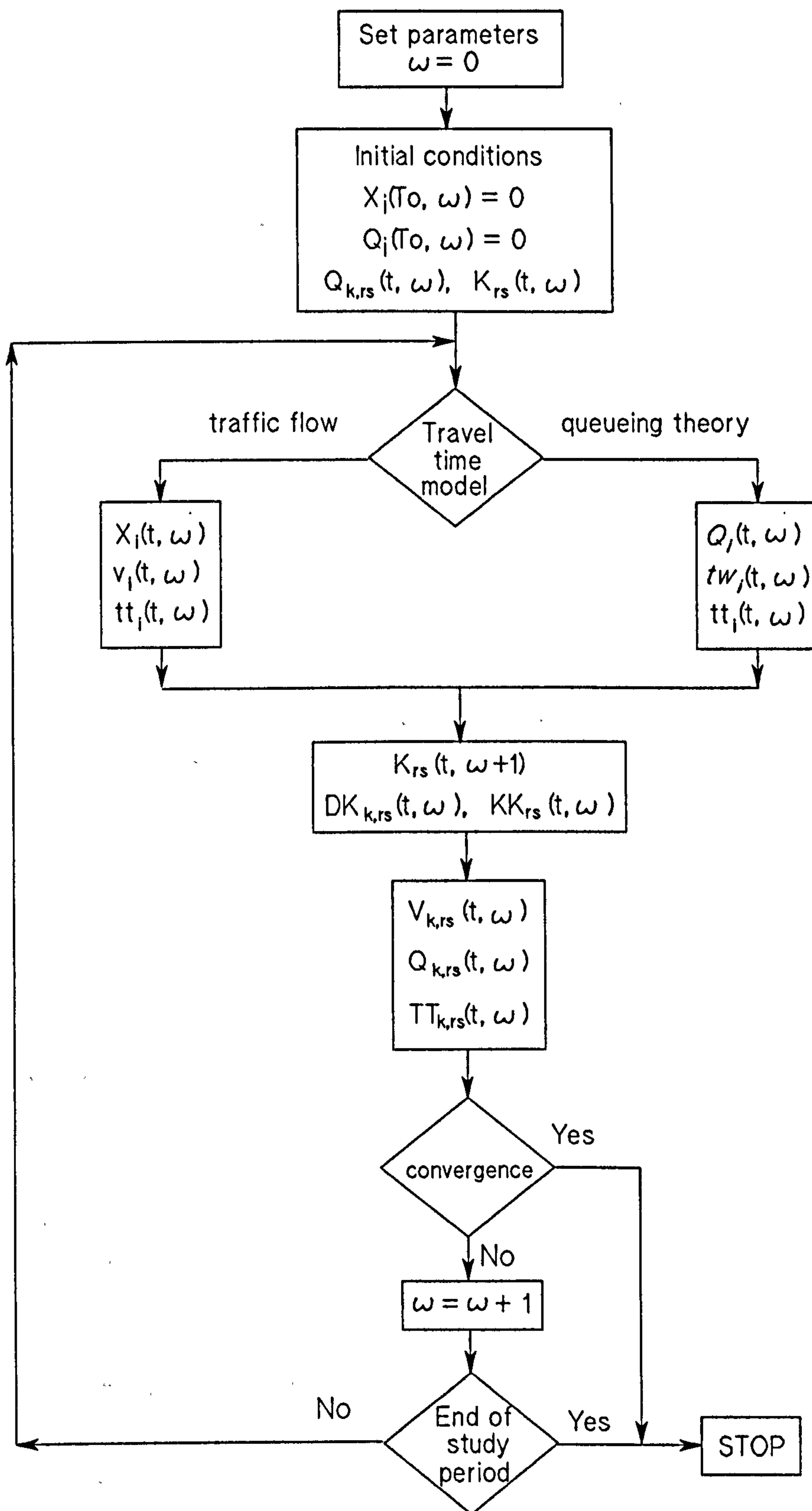


Figure 6.9 : Flowchart of the simulation algorithm.

where  $\xi$  is a predetermined tolerance factor which reflects the maximum relative deviation from the stationary state. The above convergence criterion is quite strict, (dependent on the value of  $\xi$ ), since it requires the relative deviation of  $Q_{k,rs}(t,\omega)$  to be less or equal to  $F_1 \cdot F_2 \cdot \xi$  in order the iterative process to stop.

The assumption that  $\omega$  represents a day is not restrictive, and  $\omega$  can be used to represent any time period during which trip decisions are assumed to remain constant. Furthermore, it should be noted that the general formulation of the model does not imply that there is a particular group of travellers who alter their trip choices every day. What remains constant† from day to day is the probability that a randomly selected individual will review his current trip choices. Therefore there is a constant† percentage of travellers equal to  $F_1 \cdot 100\%$  who review their current trip choices every day.

### 6.7 Summary

This chapter has developed a dynamic stochastic assignment model which can be used in order to estimate the time varying link flows and travel times in any general network. The model is useful for the analysis of the peak hour demand and deals with the situation where commuters travelling between different O-D pairs have to arrive at their destination at predetermined times.

Travellers may have the choice of arriving earlier or later than their work start time and thus to travel outside the peak and experience shorter travel times than the ones they would experience if they arrived in time. They also select one of the routes they consider as reasonable choices. A route is considered as a reasonable choice if it includes only links which do not take the traveller back towards his origin. Since the traffic conditions during the peak are not uniform, but traffic patterns vary both in space and time, the set of reasonable paths is not constant but depends on the time that a traveller decided to depart from his origin.

The model consists of three different components: a travel time model, a demand model, and a demand adjustment mechanism. The travel time model predicts the time varying traffic patterns, given the time dependent distributions of the departure rates from the origins, while the demand model estimates the latter distributions given the

---

† Subject to the condition that the route choice set remains the same on two successive days, i.e.  $K_{rs}(t,\omega) = K_{rs}(t,\omega+1) \forall t$

time dependent levels of utility. The demand adjustment mechanism models the interaction between users and network performance.

Two different approaches are used in the travel time model. They are based on traffic flow and queueing theory, and provide the framework of two macroscopic simulation models which calculate the time varying flow patterns or queue lengths, and link travel times; the latter are then used to estimate the O-D travel times.

The traffic flow theoretic approach is based on elementary relationships between speed, flow and density. Traffic conditions within a link are assumed to be homogeneous, i.e., vehicles are assumed to be uniformly distributed over the length of the link. Flow passing through a fixed point on a link is assumed to be an aggregation of different flow components corresponding to the paths which include that link. Furthermore the speed of a vehicle within a link is assumed to be constant. Traffic patterns are described by a volume conservation equation implying that the rate of change in the number of cars within a link is equal to the inflow to this link minus the outflow from the link.

A different approach to calculate O-D travel times is based in deterministic time dependent queueing theory. Each link of the network is assumed to consist of two different segments. The first has adequate capacity so that the free flow conditions are always held, and the second is a bottleneck with a fixed capacity where a queue may develop. The queue formed at the bottleneck of each link is assumed to consist of different groups of vehicles corresponding to the routes which include that link. Waiting time is a function of the queue length and the capacity of the bottleneck.

The demand model is based on the utility maximisation decision rule, and defines the time dependent departure rates following each reasonable route connecting the O-D pairs of the network, given the utility associated with each combination of departure time and route. In the demand model, an individual is assumed to first choose a departure time  $t$ , which is considered as a continuous variable, and then select a reasonable route  $k$ , conditional on the choice of  $t$ .

Travellers decide on their trip choices by trading off the travel time and schedule delay associated with each alternative combination of route and departure time, and they are assumed to select that alternative which they believe has the highest utility. However different travellers may have different perceptions of the utility derived from the same alternative, and therefore traffic patterns depend on the degree of variability of travellers' perceptions. The two dimensional decision problem faced by each individual

is modelled using a mixed continuous / discrete nested logit formulation. The utility function used, reflects the disutility due to travel time and schedule delay. O-D travel times are defined from the travel time model and are then used in order to determine schedule delays from the work starting times at destinations.

The demand adjustment mechanism is derived from a dynamic Markovian model and describes the evolution over time of the departure rate distributions. Travellers are assumed to modify their trip decisions based on the information they acquire from their recent trips. The probability that a randomly selected individual will review his current trip choices is assumed to be constant. Furthermore, travellers who review their decisions are classified in two different categories i) the ones who will switch to a different route and keep the same departure time and ii) the ones who will alter their trip decisions in both or at least in one of their dimensions of choice. The probability that a randomly selected individual who will alter his decisions, will belong to any of these two categories is constant.

Travellers' decisions on their new trip choices are based on the utility maximisation principle. Thus, the evolution of the departure rate distributions described by the demand adjustment mechanism, is derived using ;

- a multinomial logit for the travellers belonging to the first category, and
- the nested logit formulation of the demand model, for the ones belonging to the second.

The main equations of the model are modified so that continuous variables are transformed into discrete, and thus the demand adjustment mechanism represents the evolution of traffic patterns from day to day. The discrete version of the model is suitable for numerical solution and provides the framework for the computer simulation program that solves the DSUE model. The simulation algorithm is an iterative process with each iteration representing a day.

The equilibrium solution is achieved, when travellers believe that they cannot increase their utility of travel by unilaterally changing route or departure time, i.e., when the traffic patterns on two successive days (iterations) are very similar. The equilibrium patterns are therefore derived through a continuous adjustment of the departure rate distribution, performed by the demand adjustment mechanism.

# **7 formulating dynamic assignment as a mathematical program**

### **Objective**

The purpose of this chapter is to present a framework for formulating and solving the dynamic assignment problem as an equivalent optimisation program.

### **7.1. Introduction**

In the previous chapter a dynamic stochastic assignment model was developed which can handle general network forms. A main component of this model is the demand adjustment mechanism which describes the day-to-day evolution of the demand and is used in order to : (i) describe the interactions between users and network performance and, (ii) to define the equilibrium solution of the DSUE problem.

In this chapter a different method is developed. Based on the approach used in the static stochastic assignment (chapter 3), a framework for formulating and solving the DSUE problem as an equivalent optimisation program is presented. This approach defines only the equilibrium solution of the problem and it does not describe the evolution of the traffic patterns from day-to-day.

The derivation is only true under certain restrictive assumptions. Therefore the analysis should be regarded as a framework rather than as a strict formulation and further research is required to examine the implications of relaxing these assumptions. However, in terms of empirical evidence, section 8.7 demonstrates that the results derived from the demand adjustment mechanism approach (described in the previous chapter) and the equivalent program approach are identical.

The solution algorithm is based on the method of successive averages and is an iterative process. In each iteration, a dynamic stochastic network loading is performed in order to find the current solution. Following the definition of static stochastic network

loading, a *dynamic stochastic network loading* (DSNL) mechanism is a process of assigning (in time and space) a set of O-D trip rates to a transportation network in which link travel times are time varying but fixed in the sense that they are not flow dependent. This procedure is conducted assuming that route and departure time choice is based on perceived rather than measured travel times and schedule delay, and travellers select the alternative which they believe is associated with the maximum utility. Thus a form of a DSNL mechanism can be defined from the modelling procedures developed in the previous chapter, which are used in the following sequence :

- 1) Given the time varying link travel times  $tt_i(t)$ , determine the set of reasonable paths  $K_{rs}(t)$ ,  $\forall t, rs$ , as defined in section 6.2.
- 2) Calculate the O-D travel times  $TT_{k,rs}(t)$  following the procedure described in subsection 6.3.3
- 3) Calculate the utility function  $V_{k,rs}(t)$ , using equation (6.55) and then the time dependent departure rates  $Q_{k,rs}(t)$  using equation (6.56) and (6.43).
- 4) Use equation (6.15) to calculate the time dependent link volumes  $X_i(t)$  and then equation (6.9) to define the link flow patterns  $q_i(t)$ .

The DSUE assignment differs from the DSNL mechanism, since in contrast to the latter in the former, perceived travel times are modelled not only as random variables but also as flow dependent. Thus link travel times can be denoted by  $tt_i(t) = tt_i(q_i(t))$ . It should be noted that in the analysis that follows in this chapter, link travel times are assumed to be defined using the BPR volume delay curve. Therefore the time dependent link travel times are defined from the following equation :

$$tt_i(q_i(t)) = tto_i \cdot [ 1 + a \cdot (q_i(t)/c_i)^b ]$$

Consequently, the DSNL mechanism, which as will be shown is used in order to derive the DSUE traffic patterns, is based on the traffic flow theoretic approach (section 6.3.1) and the speed is defined from equation (6.22).

Given the set of O-D trip rates,  $\{ Q_{rs} \}$ , the dynamic stochastic user equilibrium conditions can be expressed as :

$$Q_{k,rs}(t) = Q_{rs} \cdot P_{k,rs}(t) \quad t \in [T_0, T_0+T], k \in K_{rs}(t), rs \quad (7.1)$$

where  $P_{k,rs}(t)$  is the probability that a route  $k$  and a departure time  $t$  is selected given the set of the measured utilities  $V_{k,rs}(t)$ . In other words

$$P_{k,rs}(t) = \Pr ( U_{k,rs}(t) \geq U_{m,rs}(t') \quad \forall (t',m) \neq (t,k), k \in K_{rs}(t), m \in K_{rs}(t') ) \quad (7.2)$$



where  $E[U_{k,rs}(t)] = V_{k,rs}(t) \quad \forall k, rs, t$

Equivalently,  $P_{k,rs}(t)$  can be expressed as :

$$P_{k,rs}(t) = \Pr (w_{k,rs}(t) < w_{m,rs}(t') \quad \forall (t',m) \neq (t,k), k \in K_{rs}(t), m \in K_{rs}(t') ) \quad (7.3)$$

where  $w_{k,rs}(t)$  is a random variable representing the perceived total disutility of travel associated with route  $k$  and departure time  $t$ , i.e.  $w_{k,rs}(t) = -U_{k,rs}(t)$ , and  $E[w_{k,rs}(t)] = W_{k,rs}(t) \quad \forall k, rs, t$ .  $W_{k,rs}(t)$  denotes the measured disutility of travel, i.e.,  $W_{k,rs}(t) = -V_{k,rs}(t)$  and can be expressed as :

$$W_{k,rs}(t) = \alpha \cdot TT_{k,rs}(t) + SD_{k,rs}(t) \quad (7.4)$$

where  $SD_{k,rs}(t)$  is the measured disutility due to schedule delay associated with a departure time  $t$  and route  $k$ .

Given the time dependent link travel times  $tt_i(q_i(t))$ , the O-D travel times  $TT_{k,rs}(t)$  can be calculated following the procedure described in subsection 6.3.3. However this variable can be also calculated in an alternative way following the notation used in static assignment. Thus let define the indicator variable  $\delta_{a,k}^{rs}(t,t')$  as :

$$\delta_{a,k}^{rs}(t,t') = \begin{cases} 1 & \text{if } k \in K_{rs}(t) \text{ includes link } a, \text{ and a vehicle departing from } r \\ & \text{at time } t, \text{ enters link } a \text{ at time } t' \\ 0 & \text{otherwise} \end{cases} \quad (7.5)$$

Then  $TT_{k,rs}(t)$  can be calculated from the following equation:

$$TT_{k,rs}(t) = \sum_a \sum_{t'} tt_a(q_a(t')) \cdot \delta_{a,k}^{rs}(t,t') \quad a \in L \quad (7.6)$$

Therefore, using equations (7.4) and (7.6) the measured disutility of travel can be expressed as:

$$W_{k,rs}(t) = -V_{k,rs}(t) = \alpha \cdot \sum_a \sum_{t'} tt_a(q_a(t')) \cdot \delta_{a,k}^{rs}(t,t') + SD_{k,rs}(t) \quad (7.7)$$

Using the indicator variable  $\delta_{a,k}^{rs}(t,t')$ , the time dependent link flow can be expressed as a function of the time dependent O-D path flows, as follows† :

$$q_a(t') = \sum_{rs} \sum_t \sum_k Q_{k,rs}(t) \cdot \delta_{a,k}^{rs}(t,t') \quad \forall a \in L, t' \in [T_0, T_0+T] \quad (7.8)$$

† In the equations used in this chapter when the summation  $\sum_t \sum_k$  is used,  $k$  is considered as  $k \in K_{rs}(t)$ .

The above equation has a form similar to equation (3.2) which express link flows in the static assignment problem. Equation (7.8) implies that the time dependent flow on each link at each point in time, is the sum of the time dependent flows on all paths and departure times, which enter that link at that point in time.

Using equation (7.1), the time dependent link flows, given by equation (7.8), can be expressed as a function of the total O-D trip rates, as follows :

$$q_a(t') = \sum_{rs} Q_{rs} \sum_t \sum_k P_{k,rs}(t) \cdot \delta_{a,k}^{rs}(t,t') \quad \forall a, t' \quad (7.9)$$

In addition the network constraints must be satisfied :

$$tt_a(t) = tt_a(q_a(t)) \quad \forall a, t \quad (7.10)$$

$$\sum_t \sum_k Q_{k,rs}(t) = Q_{rs} \quad (7.11)$$

The choice probabilities  $P_{k,rs}(t)$  are dependent on the measured levels of utilities  $V_{k,rs}(t)$ . If these utilities are known, the departure rate patterns  $Q_{k,rs}(t)$  and the link flow patterns  $q_a(t')$  which solve equations (7.8) and (7.9), respectively, can be derived by a dynamic stochastic network loading. Furthermore, it should be noted that condition (7.11) is automatically satisfied if the time dependent path flows are given by equation (7.1), since

$$\sum_t \sum_k Q_{k,rs}(t) = \sum_t \sum_k Q_{rs} \cdot P_{k,rs}(t) = Q_{rs} \sum_t \sum_k P_{k,rs}(t) = Q_{rs} \quad (7.12)$$

Equation (7.1) characterises the dynamic stochastic user equilibrium conditions defined in sections 5.3.2 and 6.1. At DSUE no motorist can improve his perceived total utility of travel by unilaterally changing routes and/or departure times, since the probability of selecting a particular route  $k$  and departing at a specific time  $t$ , is the probability that the perceived disutility of the alternative selected is lower than the perceived disutility of all the other reasonable alternatives. At the DSUE state, the equilibrium levels of utility  $V_{k,rs}(t)$  associated with each reasonable combination of route and departure time will be such that equation (7.1) is satisfied for the equilibrium path flows  $Q_{k,rs}(t)$ . These path flows, in turn, will be associated with link flows  $q_a(t)$  which satisfy equation (7.8) and which when used in equation (7.10) will result in the equilibrium time dependent link travel times  $tt_a(q_a(t))$ . The latter when used in equation (7.7) will provide the equilibrium utility levels  $V_{k,rs}(t)$ .

Having defined the mathematical formulation of the DSUE conditions, and the framework for a DSNL mechanism, the following section will formulate the DSUE equivalent minimisation program. Section 7.3 proves the equivalence between the

solution of this program and the DSUE equations, while in section 7.4 the uniqueness of its solution is demonstrated. Section 7.5 describes an algorithmic solution approach to this problem, and section 7.6 summarises the chapter.

## 7.2 The DSUE program formulation

As in the case of the static stochastic equilibrium assignment, in this section the DSUE assignment problem is formulated as a minimisation program, the solution of which is the desired time dependent equilibrium flows. The objective function to be minimised has a form similar to the one used in the static stochastic assignment problem (equation (3.29)). The proposed function, is modified to incorporate the time dependent conditions occurring in an urban transportation network; it further includes the expected total disutility of travel, instead of the total travel time as used in equation (3.29), and also has an additional term to incorporate the disutility due to schedule delay.

Thus consider the following minimisation problem :

$$\begin{aligned} \min_{\mathbf{q}} z(\mathbf{q}) = & - \sum_{rs} Q_{rs} E \left[ \min_{\substack{k \in K_{rs}(t) \\ t \in [T_o, T_o+T]}} \{w_{k,rs}(t)\} \mid W_{rs}(t) \right] + \sum_{rs} \sum_t \sum_k Q_{k,rs}(t) \cdot SD_{k,rs}(t) \\ & + \alpha \cdot \sum_t \sum_a q_a(t) \cdot tt_a(q_a(t)) - \alpha \cdot \sum_t \sum_a \int_0^{q_a(t)} tt_a(u) du \end{aligned} \quad (7.13)$$

where  $W_{rs}(t) = ( \dots, W_{k,rs}(t_1), W_{k,rs}(t_2), \dots, W_{k,rs}(t_n), \dots )$ ,  $k \in K_{rs}$ , is the vector of the time dependent measured disutility of travel†.

In the following sections of this chapter, the time dependent flow pattern that minimises equation (7.13) is proved to satisfy the DSUE conditions. Furthermore at the solution point, the network constraints are satisfied and consequently equation (7.13) can be minimised as an unconstrained program.

The objective function (7.13) in its first term includes the expected perceived disutility function, discussed in section 2.8 (eq. 2.77). Thus the program (7.13) can be rewritten as :

† Time  $t$  may then take any value  $t_1, t_2, \dots, t_n$ , such that  $t_1 < t_2 < \dots < t_n$ , where  $t_1$  corresponds to  $T_o$  and  $t_n$  corresponds to  $T_o+T$ .

$$\begin{aligned} \min_{\mathbf{q}} z(\mathbf{q}) = & - \sum_{rs} Q_{rs} \cdot S_{rs}[W_{rs}(t)] + \sum_{rs} \sum_t \sum_k Q_{k,rs}(t) \cdot SD_{k,rs}(t) \\ & + \alpha \cdot \sum_t \sum_a q_a(t) \cdot tt_a(q_a(t)) - \alpha \cdot \sum_t \sum_a \int_0^{q_a(t)} tt_a(u) du \end{aligned} \quad (7.14)$$

where  $S_{rs}[W_{rs}(t)]$  is the expected minimum perceived disutility function, and is expressed as :

$$S_{rs}[W_{rs}(t)] = E \left[ \min_{\substack{k \in K_{rs}(t) \\ t \in [T_0, T_0+T]}} \{w_{k,rs}(t)\} \mid W_{rs}(t) \right] \quad (7.15)$$

The conditioning of the random variable  $w_{k,rs}(t)$  on  $W_{rs}(t)$  in the optimisation program (7.13), implies that the expectation is taken at a given flow level  $\mathbf{q}(t)$ †. In the analysis that follows in this chapter, two of the properties of the expected perceived travel time function (section 2.8) will be used. First, this function is concave with respect to  $W_{rs}$ , and second

$$\frac{\partial S_{rs}[W_{rs}(t)]}{\partial W_{k,rs}(t)} = P_{k,rs}(t) \quad (7.16)$$

### 7.3 Equivalence conditions

In order to show that the solution of the program given by equation (7.14) and the dynamic stochastic user equilibrium equations, are equivalent, the first order conditions of this program have to coincide with the DSUE conditions.

Since the objective function to be minimised, has no constraints, the first-order condition for a minimum at the equilibrium flow pattern  $\mathbf{q} = \mathbf{q}(t)$ , is that the gradient vanish at  $\mathbf{q}$ . The gradient is taken with respect to  $\mathbf{q}$ , the vector of the time dependent link flows, and therefore the first order condition is expressed :

$$\nabla z(\mathbf{q}) = \left( \frac{\partial z(\mathbf{q})}{\partial q_1(t_1)}, \frac{\partial z(\mathbf{q})}{\partial q_1(t_2)}, \dots, \frac{\partial z(\mathbf{q})}{\partial q_1(t_n)}, \dots, \frac{\partial z(\mathbf{q})}{\partial q_\ell(t_1)}, \dots, \frac{\partial z(\mathbf{q})}{\partial q_\ell(t_n)} \right)$$

---

†  $\mathbf{q}(t)$  is the vector of the time dependent link flows defined as :

$$\mathbf{q}(t) = (q_1(t_1), q_1(t_2), \dots, q_1(t_n), \dots, q_\ell(t_1), \dots, q_\ell(t_n))$$

where  $\ell$  is the number of links in the network.

meaning that at the minimum point, each component of the gradient has to be equal to zero. In other words :

$$\frac{\partial z(q)}{\partial q_b(t')} = 0 \quad \forall b = 1, 2, \dots, \ell \quad t' = t_1, t_2, \dots, t_n \quad (7.17)$$

Thus, below, the gradient  $\nabla z(q)$  will be derived, by concentrating on the typical term  $\partial z(q)/\partial q_b(t')$ , i.e., the partial derivative of  $z(q)$  with respect to the flow on link  $b$  at time  $t'$ .

The objective function (7.14) consists of four separate summation terms. The derivative of the first term with respect to  $q_b(t')$  can be calculated as follows :

$$\frac{\partial}{\partial q_b(t')} \left[ - \sum_{rs} Q_{rs} S_{rs}[W_{rs}(t)] \right] = - \sum_{rs} Q_{rs} \sum_t \sum_k \frac{\partial S_{rs}[W_{rs}(t)]}{\partial W_{k,rs}(t)} \cdot \frac{\partial W_{k,rs}(t)}{\partial q_b(t')} \quad (7.18)$$

However in this expression

$$\frac{\partial S_{rs}[W_{rs}(t)]}{\partial W_{k,rs}(t)} = P_{k,rs}(t) \quad (7.19)$$

Furthermore using equation (7.4) the derivative of the disutility with respect to link flow can be expressed as :

$$\frac{\partial W_{k,rs}(t)}{\partial q_b(t')} = \frac{\partial [\alpha TT_{k,rs}(t) + SD_{k,rs}(t)]}{\partial q_b(t')} = \frac{\partial \alpha TT_{k,rs}(t)}{\partial q_b(t')} + \frac{\partial SD_{k,rs}(t)}{\partial q_b(t')} \quad (7.20)$$

Following equation (7.6), and using the assumption that the variable  $\delta_{a,k}^{rs}(t,t'')$  is independent of the traffic flow patterns, i.e., that  $d\delta_{a,k}^{rs}(t,t'')/dq_b(t') = 0 \quad \forall a,b,k,t,t'',t''$ , then the derivative of the O-D travel time,  $TT_{k,rs}(t)$  with respect to link flow  $q_b(t')$  is calculated as :

$$\frac{\partial TT_{k,rs}(t)}{\partial q_b(t')} = \frac{\partial}{\partial q_b(t')} \left[ \sum_{t''} \sum_a tt_a(q_a(t'')) \cdot \delta_{a,k}^{rs}(t,t'') \right] = \frac{dtt_b(q_b(t'))}{dq_b(t')} \cdot \delta_{b,k}^{rs}(t,t') \quad (7.21)$$

Therefore, using equation (7.19), (7.20) and (7.21) in equation (7.18) the latter is expressed as :

$$\frac{\partial}{\partial q_b(t')} \left[ - \sum_{rs} Q_{rs} S_{rs}[W_{rs}(t)] \right] =$$

$$= - \sum_{rs} Q_{rs} \sum_t \sum_k P_{k,rs}(t) \cdot \left[ \frac{\alpha \cdot dtt_b(q_b(t'))}{dq_b(t')} \delta_{b,k}^{rs}(t,t') + \frac{\partial SD_{k,rs}(t)}{\partial q_b(t')} \right] \quad (7.22)$$

The second term in the objective function represents the total measured disutility due to schedule delay. The derivative of this term with respect to  $q_b(t')$  is expressed as :

$$\sum_{rs} \sum_t \sum_k Q_{k,rs}(t) \cdot \frac{\partial SD_{k,rs}(t)}{\partial q_b(t')} \quad (7.23)$$

The third term represents the total disutility due to travel time. The derivative of this term with respect to  $q_b(t')$  is given by :

$$\frac{\partial}{\partial q_b(t')} \left[ \alpha \cdot \sum_t \sum_a q_a(t) \cdot tt_a(q_a(t)) \right] = \alpha \cdot tt_b(q_b(t')) + q_b(t') \frac{\alpha \cdot dtt_b(q_b(t'))}{dq_b(t')} \quad (7.24)$$

and the derivative of the last term is :

$$\frac{\partial}{\partial q_b(t')} \left[ - \alpha \cdot \sum_t \sum_a \int_0^{q_a(t)} tt_a(u) du \right] = - \alpha \cdot tt_b(q_b(t')) \quad (7.25)$$

Thus using equations (7.22), (7.23), (7.24) and (7.25), a typical component of the gradient of the objective function, represented by the derivative of  $z(q)$  with respect to  $q_b(t')$  is expressed as :

$$\begin{aligned} \frac{\partial z(q)}{\partial q_b(t')} &= - \sum_{rs} Q_{rs} \sum_t \sum_k P_{k,rs}(t) \frac{\alpha \cdot dtt_b(q_b(t'))}{dq_b(t')} \delta_{b,k}^{rs}(t,t') - \\ &\quad - \sum_{rs} \sum_t \sum_k Q_{k,rs}(t) \cdot \frac{\partial SD_{k,rs}(t)}{\partial q_b(t')} + \sum_{rs} \sum_t \sum_k Q_{k,rs}(t) \cdot \frac{\partial SD_{k,rs}(t)}{\partial q_b(t')} \\ &\quad + \alpha \cdot tt_b(q_b(t')) + q_b(t') \frac{\alpha \cdot dtt_b(q_b(t'))}{dq_b(t')} - \alpha \cdot tt_b(q_b(t')) = \\ &= \frac{\alpha \cdot dtt_b(q_b(t'))}{dq_b(t')} \left[ q_b(t') - \sum_{rs} Q_{rs} \sum_t \sum_k P_{k,rs}(t) \delta_{b,k}^{rs}(t,t') \right] \quad (7.26) \end{aligned}$$

Assuming that the link performance functions are strictly increasing functions of flow (i.e.  $dt_b(q_b(t'))/dq_b(t') > 0$ ), and given that  $\alpha > 0$ , the gradient will vanish only when :

$$q_b(t') = \sum_{rs} Q_{rs} \sum_t \sum_k P_{k,rs}(t) \delta_{b,k}^{rs}(t,t') \quad (7.27)$$

Therefore, based on equation (7.26), the gradient vector of  $z(q)$  can be written explicitly, as :

$$\nabla z(q) = \alpha \cdot \left[ -\sum_{rs} Q_{rs} \cdot P_{rs} \cdot (\Delta_{rs})^T + q \right] \cdot \nabla_q tt \quad (7.28)$$

where :

$P_{rs}$  is the vector of the time dependent path choice probabilities, expressed as

$$P_{rs} = ( \dots, P_{k,rs}(t_1), P_{k,rs}(t_2), \dots, P_{k,rs}(t_n), \dots ) \quad \forall k \in K_{rs}$$

$\nabla_q tt$  is the Jacobian matrix† of the time dependent link travel times. This matrix includes the derivative of each link performance function with respect to each time dependent link flow and will be analysed in the following section.

$\Delta_{rs}$  is the link-entry time — path-departure time incidence matrix.

It is an  $\ell.n \times K_{rs}.n$  matrix, where  $\ell$  is the number of links in the network,  $K_{rs}$  is the number of elements in the set  $K_{rs}$ , i.e., defines the number of reasonable paths connecting the O-D pair r-s during the period of analysis, and n determines the length of the latter period, since  $t_n$  is used in order to

† The Jacobian of a vector of functions  $f(x) = ( \dots f_a(x), \dots )$ , includes the partial derivatives of the function  $f_a(x)$  with respect to all the values  $x_i$  and is expressed as:

$$\nabla_x f = \begin{vmatrix} \frac{\partial f_1(x)}{\partial x_1} & \frac{\partial f_2(x)}{\partial x_1} & \dots & \frac{\partial f_a(x)}{\partial x_1} & \dots \\ \frac{\partial f_1(x)}{\partial x_2} & \frac{\partial f_2(x)}{\partial x_2} & \dots & \frac{\partial f_a(x)}{\partial x_2} & \dots \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \frac{\partial f_1(x)}{\partial x_i} & \frac{\partial f_2(x)}{\partial x_i} & \dots & \frac{\partial f_a(x)}{\partial x_i} & \dots \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \end{vmatrix}$$

express the time  $T_0+T$ . Thus the element in the  $(a-1).n + i$  row and  $(k-1).n + j$  column is the element  $\delta_{a,k}^{rs}(t_j, t_i)$ .

Equation (7.27) expresses the equilibrium time dependent link flows corresponding to the equilibrium time dependent path flows given by equation (7.1). This can be verified by multiplying both sides of equation (7.1) by  $\delta_{b,k}^{rs}(t, t')$  and summing over all departure times  $t$  and paths  $k \in K_{rs}(t)$  connecting all O-D pairs  $r-s$  :

$$(7.1) \Leftrightarrow Q_{k,rs}(t) = Q_{rs} \cdot P_{k,rs}(t) \Rightarrow$$

$$\Rightarrow \sum_{rs} \sum_t \sum_k Q_{k,rs}(t) \cdot \delta_{b,k}^{rs}(t, t') = \sum_{rs} \sum_t \sum_k Q_{rs} \cdot P_{k,rs}(t) \cdot \delta_{b,k}^{rs}(t, t') \Rightarrow$$

$$\stackrel{\text{eq. (7.8)}}{\Rightarrow} q_b(t') = \sum_{rs} Q_{rs} \sum_t \sum_k P_{k,rs}(t) \cdot \delta_{b,k}^{rs}(t, t')$$

Furthermore, since  $\sum_t \sum_k P_{k,rs}(t) = 1$ , both sides of equation (7.1) can be summed over all departure times and all paths connecting  $r-s$  and result in the condition :

$$\sum_t \sum_k Q_{k,rs}(t) = Q_{rs}$$

Thus the flow pattern that solves the DSUE program also satisfies the DSUE conditions and the flow conservation constraints. Consequently this flow pattern can be obtained by minimising the DSUE program.

#### 7.4 Uniqueness Conditions

In the previous section, the solution of the minimisation program defined by equation (7.13) or (7.14), was proved to be equivalent to the DSUE conditions. However in order the equivalent minimisation program formulation to be useful the uniqueness of its solution must be proved. To do that it is sufficient to show that the objective function (7.13) is strictly convex in the vicinity of the minimum solution, and convex elsewhere. A sufficient condition for proving the convexity of the objective function, is to show that the Hessian matrix of the DSUE objective function is positive definite. This matrix will be calculated below, by focusing on the representative term,  $\partial^2 z(q) / \partial q_b(t') \partial q_a(t'')$ .

The first derivative of the objective function is given by equation (7.26). Using this equation the second derivative of  $z(q)$  can be expressed as :



$$\frac{\partial z^2(q)}{\partial q_b(t') \partial q_a(t'')} = \left[ q_b(t') - \sum_{rs} Q_{rs} \sum_t \sum_k P_{k,rs}(t) \delta_{b,k}^{rs}(t,t') \right] \frac{\alpha \cdot d^2 tt_b(q_b(t'))}{dq_b(t') dq_a(t'')} +$$

$$+ \left[ \frac{\partial q_b(t')}{\partial q_a(t'')} - \sum_{rs} \sum_t \sum_k Q_{rs} \sum_i \sum_{t_i} \frac{\partial P_{k,rs}(t)}{\partial W_{i,rs}(t_i)} \cdot \frac{\partial W_{i,rs}(t_i)}{\partial q_a(t'')} \delta_{b,k}^{rs}(t,t') \right] \frac{\alpha \cdot dtt_b(q_b(t'))}{dq_b(t')} \quad (7.29)$$

However, using equation (7.4) and following the procedure for the utility function specification (section 6.4.1) the disutility of travel can be expressed as:

$$W_{i,rs}(t) = \begin{cases} \alpha \cdot TT_{i,rs}(t) + \beta \cdot [ (t_{rs} - D_{rs}) - (t + TT_{i,rs}(t)) ] & \text{for early arrivals} \\ \alpha \cdot TT_{i,rs}(t) & \text{for on time arrivals} \\ \alpha \cdot TT_{i,rs}(t) + \beta \cdot \gamma \cdot [ (t + TT_{i,rs}(t)) - (t_{rs} + D_{rs}) ] & \text{for late arrivals} \end{cases}$$

and therefore

$$\frac{\partial W_{i,rs}(t)}{\partial TT_{i,rs}(t)} = \begin{cases} \alpha - \beta & \text{for early arrivals} \\ \alpha & \text{for on time arrivals} \\ \alpha + \beta \cdot \gamma & \text{for late arrivals} \end{cases} \quad (7.30)$$

Assuming that  $\alpha > \beta$ , in other words that travel time is more onerous than early schedule delay, (which as was discussed in section 4.3 is a reasonable assumption), then the following relationship is satisfied :

$$\frac{\partial W_{i,rs}(t)}{\partial TT_{i,rs}(t)} > 0 \quad \forall rs, t=t_1, t_2, \dots, t_n, i \in K_{rs}(t) \quad (7.31)$$

Furthermore ..

$$\frac{\partial W_{i,rs}(t)}{\partial q_a(t'')} = \frac{\partial W_{i,rs}(t)}{\partial TT_{i,rs}(t)} \cdot \frac{\partial TT_{i,rs}(t)}{\partial q_a(t'')} \quad (7.32)$$

Thus using the above equation and equation (7.21), the derivative of the disutility of travel with respect to link flow can be expressed as:

$$\frac{\partial W_{i,rs}(t)}{\partial q_a(t'')} = \frac{\partial W_{i,rs}(t)}{\partial TT_{i,rs}(t)} \cdot \frac{dtt_a(q_a(t''))}{dq_a(t'')} \cdot \delta_{a,i}^{rs}(t,t'') \quad (7.33)$$

In addition to the above conditions,

$$\frac{\partial q_b(t')}{\partial q_a(t'')} = \begin{cases} 1 & \text{if } a = b \text{ and } t' = t'' \\ 0 & \text{otherwise} \end{cases} \quad (7.34)$$

Substituting (7.33) and (7.34) into equation (7.29) the second derivative of  $z(q)$  is expressed as :

$$\begin{aligned} \frac{\partial^2 z(q)}{\partial q_b(t') \partial q_a(t'')} = & \left\{ \begin{aligned} & \left[ q_b(t') - \sum_{rs} Q_{rs} \sum_t \sum_k P_{k,rs}(t) \delta_{b,k}^{rs}(t,t') \right] \frac{\alpha \cdot d^2 tt_b(q_b(t'))}{dq_b(t')^2} - \\ & - \sum_{rs} Q_{rs} \sum_k \sum_i \sum_t \sum_{t_i} \frac{\partial P_{k,rs}(t)}{\partial W_{i,rs}(t_i)} \cdot \frac{\partial W_{i,rs}(t_i)}{\partial \Gamma_{i,rs}(t_i)} \cdot \left[ \frac{dtt_a(q_a(t''))}{dq_a(t'')} \delta_{a,i}^{rs}(t_i,t'') \right] \\ & \cdot \left[ \frac{\alpha \cdot dtt_b(q_b(t'))}{dq_b(t')} \delta_{b,k}^{rs}(t,t') \right] + \frac{\alpha \cdot dtt_b(q_b(t'))}{dq_b(t')} \quad \text{for } (a,t') = (b,t'') \\ & \hspace{15em} (7.35a) \end{aligned} \right. \\ = & \left\{ \begin{aligned} & - \sum_{rs} Q_{rs} \sum_k \sum_i \sum_t \sum_{t_i} \frac{\partial P_{k,rs}(t)}{\partial W_{i,rs}(t_i)} \cdot \frac{\partial W_{i,rs}(t_i)}{\partial \Gamma_{i,rs}(t_i)} \cdot \left[ \frac{dtt_a(q_a(t''))}{dq_a(t'')} \delta_{a,i}^{rs}(t_i,t'') \right] \\ & \cdot \left[ \frac{\alpha \cdot dtt_b(q_b(t'))}{dq_b(t')} \delta_{b,k}^{rs}(t,t') \right] \quad \text{for } (a,t') \neq (b,t'') \\ & \hspace{15em} (7.35b) \end{aligned} \right. \end{aligned}$$

The above equation gives the term in the  $(b-1).n+t'$  row and the  $(a-1).n+t''$  column of the Hessian matrix of  $z(q)$ . The Hessian can be expressed as the sum of three separate matrices as follows :

$$\begin{aligned} \nabla^2 z(q) = & \alpha \cdot \nabla_q^2 tt.R + \alpha \cdot \sum_{rs} Q_{rs} \cdot [ (\nabla_q tt \cdot \Delta_{rs}) \cdot [ (\nabla_{TT} W_{rs}) \cdot (-\nabla_W P_{rs}) ] \cdot \\ & \cdot (\nabla_q tt \cdot \Delta_{rs})^T ] + \alpha \cdot \nabla_q tt \end{aligned} \quad (7.36)$$

The Hessian is an  $(\ell \cdot n) \times (\ell \cdot n)$  matrix, where  $\ell$  is the number of links in the network, and  $n$  defines the duration of the study period. The notations used in the above equation are described below :

- (i)  $\nabla_q tt$  is the  $(\ell \cdot n) \times (\ell \cdot n)$  Jacobian of the time dependent link travel time vector, having the following form

$$\nabla_{q} tt = \begin{vmatrix} \nabla_{q_1} tt_1 & \nabla_{q_1} tt_2 & \nabla_{q_1} tt_3 & \dots & \nabla_{q_1} tt_\ell \\ \nabla_{q_2} tt_1 & \nabla_{q_2} tt_2 & \nabla_{q_2} tt_3 & \dots & \nabla_{q_2} tt_\ell \\ \nabla_{q_3} tt_1 & \nabla_{q_3} tt_2 & \nabla_{q_3} tt_3 & \dots & \nabla_{q_3} tt_\ell \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \nabla_{q_p} tt_1 & \nabla_{q_p} tt_2 & \nabla_{q_p} tt_3 & \dots & \nabla_{q_p} tt_\ell \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \nabla_{q_\ell} tt_1 & \nabla_{q_\ell} tt_2 & \nabla_{q_\ell} tt_3 & \dots & \nabla_{q_\ell} tt_\ell \end{vmatrix}$$

where :

$$\nabla_{q_p} tt_s = \begin{vmatrix} \frac{\partial tt_s(q_s(t_1))}{\partial q_p(t_1)} & \frac{\partial tt_s(q_s(t_2))}{\partial q_p(t_1)} & \dots & \frac{\partial tt_s(q_s(t_n))}{\partial q_p(t_1)} \\ \frac{\partial tt_s(q_s(t_1))}{\partial q_p(t_2)} & \frac{\partial tt_s(q_s(t_2))}{\partial q_p(t_2)} & \dots & \frac{\partial tt_s(q_s(t_n))}{\partial q_p(t_2)} \\ \vdots & \vdots & \dots & \vdots \\ \frac{\partial tt_s(q_s(t_1))}{\partial q_p(t_n)} & \frac{\partial tt_s(q_s(t_2))}{\partial q_p(t_n)} & \dots & \frac{\partial tt_s(q_s(t_n))}{\partial q_p(t_n)} \end{vmatrix}$$

In the case examined, it is assumed that there are no link interactions, i.e., the travel time on a link depends only on the flow that traverse this link, and thus :

$$\nabla_{q_p} tt_s = \begin{vmatrix} 0 & 0 & 0 & \dots & 0 \\ 0 & 0 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \dots & \vdots \\ \cdot & \cdot & \cdot & \dots & \cdot \\ \vdots & \vdots & \vdots & \dots & \vdots \\ \cdot & \cdot & \cdot & \dots & \cdot \\ 0 & 0 & 0 & \dots & 0 \end{vmatrix} \quad \text{for } p \neq s$$



(ii)  $\nabla_q^2 tt$  is a diagonal  $\ell.n \times \ell.n$  matrix which includes elements  $d^2 tt_a(q_a(t_i)) / dq_a(t_i)^2$  along its diagonal.

(iii)  $\Delta_{rs}$  is the  $\ell.n \times K_{rs}.n$  link-entry time - path-departure time incidence matrix, defined in section 7.2.

(iv) The matrix, R, is a diagonal  $\ell.n \times \ell.n$  matrix which includes terms of the form :

$$q_b(t') - \sum_{rs} Q_{rs} \sum_t \sum_k P_{k,rs}(t) \delta_{b,k}^{rs}(t,t')$$

the above is the  $(b-1).n + t'$  element along the diagonal

(v)  $\nabla_W P_{rs}$  is the Jacobian of the choice probability vector  $P_{rs}$  for the O-D pair r-s, with respect to the disutilities of travel  $W_{rs}(t)$ . It is a  $K_{rs}.n \times K_{rs}.n$  matrix. The element on the  $(m-1).n + t'$  row and  $(k-1).n + t$  column is of the form :

$$\partial P_{k,rs}(t) / \partial W_{m,rs}(t')$$

(vi)  $(\nabla_{TT} W_{rs})$  is a diagonal  $K_{rs}.n \times K_{rs}.n$  matrix which includes elements of the form

$$\partial W_{k,rs}(t) / \partial TT_{k,rs}(t) \text{ along its diagonal.}$$

The first matrix in the sum that comprises the Hessian (eq. 7.36) is the product of two different matrices :

- The first one,  $\nabla_q^2 tt$ , is a diagonal matrix. Furthermore, assuming that the link performance functions are convex, the elements of this matrix, are positive. Therefore, since  $\nabla_q^2 tt$  is a diagonal with positive entries, it is a positive definite matrix.

- The second matrix, R, in the product is also a diagonal matrix which includes terms of the form  $q_b(t') - \sum_{rs} Q_{rs} \sum_t \sum_k P_{k,rs}(t) \delta_{b,k}^{rs}(t,t')$  along its diagonal. However these terms can be either positive or negative. Thus the first matrix in equation (7.36), is an indefinite matrix.

The second matrix in the right hand side of equation (7.35) includes a quadratic form applied to the matrix  $[(\nabla_{TT} W_{rs}).(-\nabla_W P_{rs})]$ , and is expressed as :

$$\alpha \cdot \sum_{rs} Q_{rs} [ (\nabla_q tt \cdot \Delta^{rs}) \cdot [(\nabla_{TT} W_{rs}).(-\nabla_W P_{rs})] \cdot (\nabla_q tt \cdot \Delta^{rs})^T ] \quad (7.37)$$

The matrix  $\nabla_{TT} W_{rs}$  is a diagonal with positive entries, assuming that the disutility (per unit time) due to travel time is higher than the one due to early schedule delay, (eq. (7.31)). It is therefore a positive definite matrix. The Jacobian  $\nabla_W P_{rs}$  is also the

Hessian of the expected disutility of travel†,  $S_{rs}(W_{rs})$ , which as was discussed in section 2.8, is concave in  $W_{rs}$ . Therefore this matrix is negative semidefinite, implying that  $(-\nabla_w P_{rs})$  is positive semidefinite. The product of the positive definite diagonal matrix  $\nabla_{TT} W_{rs}$ , and the positive semidefinite  $\nabla_w P_{rs}$  is a positive semidefinite matrix. Equation (7.37) includes then a quadratic form applied to a positive semidefinite matrix; therefore is positive semidefinite.

The third matrix in equation (7.36) is  $\alpha \cdot \nabla_q tt$ . This is a diagonal matrix with positive entries, since the link performance functions are assumed to be convex. It is therefore a positive definite matrix.

The matrix representing the sum of the second and the third matrix in the sum that comprises the Hessian, is therefore positive definite, since it is the sum of a positive semidefinite and a positive definite matrix. However, the Hessian includes also the indefinite matrix  $R$  and is therefore an indefinite matrix, i.e. it can be either positive definite or negative definite.

However it should be noted that in the vicinity of the equilibrium point, and when this point is approached, the gradient of the objective function tends to zero, and thus  $(q_b(t') - \sum_{rs} Q_{rs} \sum_t \sum_k P_{k,rs}(t) \delta_{b,k}^{rs}(t,t')) \rightarrow 0$ , and the first matrix in equation (7.36) vanishes. This means that at the minimum point that satisfies the DSUE conditions, the Hessian of the objective function is a positive definite matrix, and thus the DSUE equivalent program is strictly convex. However at all the other points the Hessian of  $z(q)$  is an indefinite matrix, since it is the sum of a positive definite and an indefinite matrix.

The only conclusion that can be drawn from the above analysis is that the DSUE equilibrium flow pattern is a local minimum of the equivalent objective function (7.13). No conclusion can be drawn on whether this point is the global minimum, or whether there are some other local minima in the equivalent DSUE minimisation program.

In order to show that the DSUE point is the global minimum of the DSUE equivalent program, the simple transformation of the performance function, which was suggested by Sheffi and Powell (1982) in the analysis of static stochastic assignment, will be employed. Thus, since  $tt_a(q_a(t))$  is a monotone, an inversion of this function is performed, where  $q_a(tt_a(t))$  represents the inverse of  $tt_a(q_a(t))$ .

† Because  $P_{k,rs}(t)$  is derived from equation (7.16).

Figure 7.1a and b illustrates a link performance function and its inverse, respectively. The inverse function  $q_a(tt_a(t))$  is increasing for all positive  $tt_a(t)$ , and positive for all  $tt_a(t) > tt_a(0)$ .

Thus the DSUE program can be transformed to a function of  $tt$  rather than of  $q$ , by introducing the change of variables  $q_a(t)=q_a(tt_a(t))$  in equation (7.13). The objective function, then, becomes :

$$z(tt) = - \sum_{rs} Q_{rs} E \left[ \min_{\substack{k \in K_{rs}(t) \\ t \in [T_0, T_0+T]}} \{W_{k,rs}(t)\} \mid W_{rs}(t) \right] + \sum_{rs} \sum_t \sum_k Q_{k,rs}(t) \cdot SD_{k,rs}(t) + \\ + \alpha \cdot \sum_t \sum_a q_a(tt_a(t)) \cdot tt_a(t) - \alpha \cdot \sum_t \sum_a \int_{tt_a(0)}^{tt_a(t)} u \frac{dq_a(u)}{du} du \quad (7.38)$$

Thus, the new optimisation program, after integrating  $z(tt)$  by parts, becomes :

$$\min_{tt} z(tt) = - \sum_{rs} Q_{rs} E \left[ \min_{\substack{k \in K_{rs}(t) \\ t \in [T_0, T_0+T]}} \{W_{k,rs}(t)\} \mid W_{rs}(t) \right] + \sum_{rs} \sum_t \sum_k Q_{k,rs}(t) \cdot SD_{k,rs}(t) + \\ + \alpha \cdot \sum_t \sum_a \int_{tt_a(0)}^{tt_a(t)} q_a(u) du \quad (7.39)$$

The first derivative of  $z(tt)$  with respect to  $tt_b(t')$  can then be derived, following the same procedure as in the calculation of the  $z(q)/dq_b(t')$ . Thus :

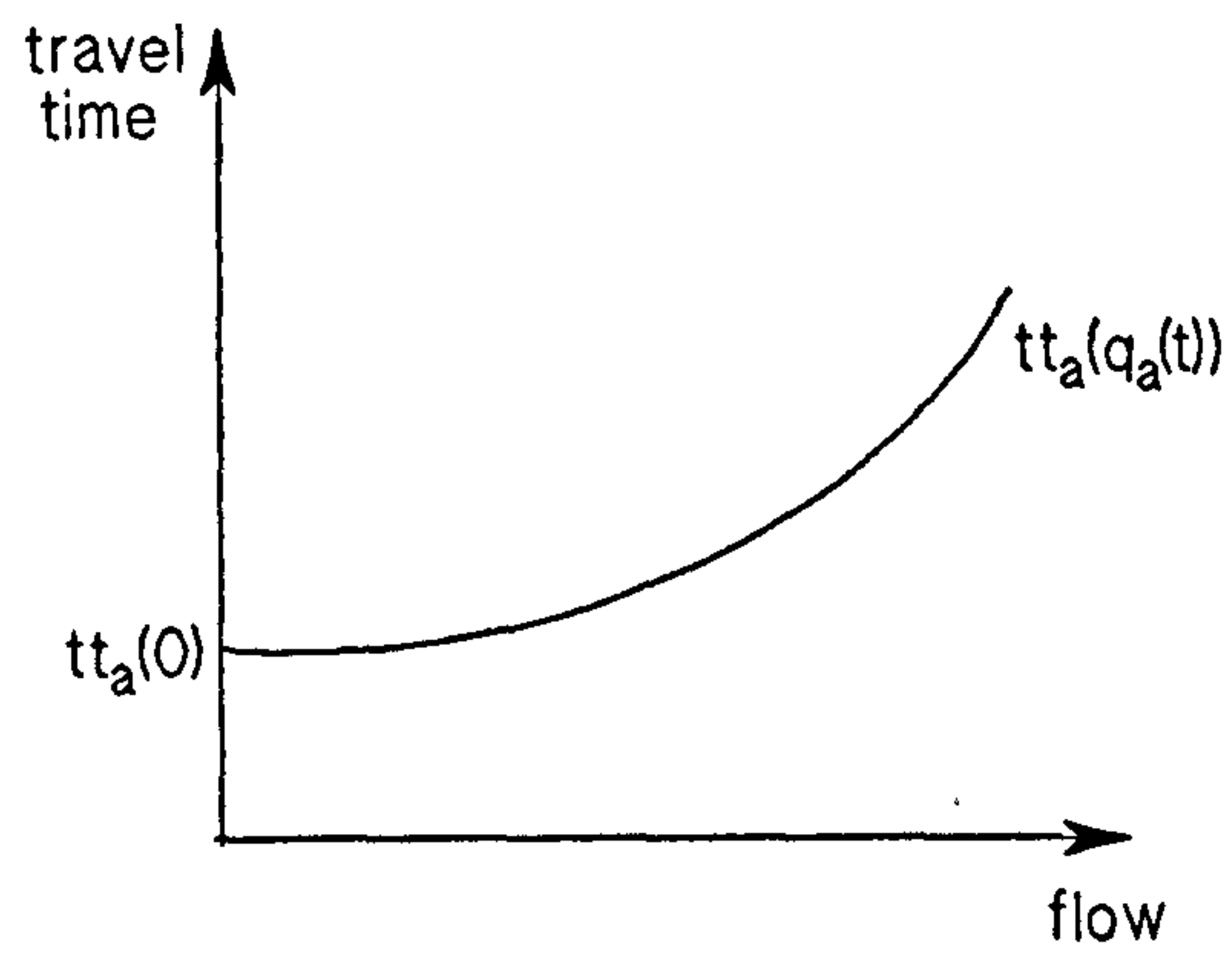
$$\frac{\partial TT_{k,rs}(t)}{\partial tt_b(t')} = \delta_{b,k}^{rs}(t,t') \quad (7.40)$$

and therefore, using (7.18), (7.19), (7.20) and (7.40) :

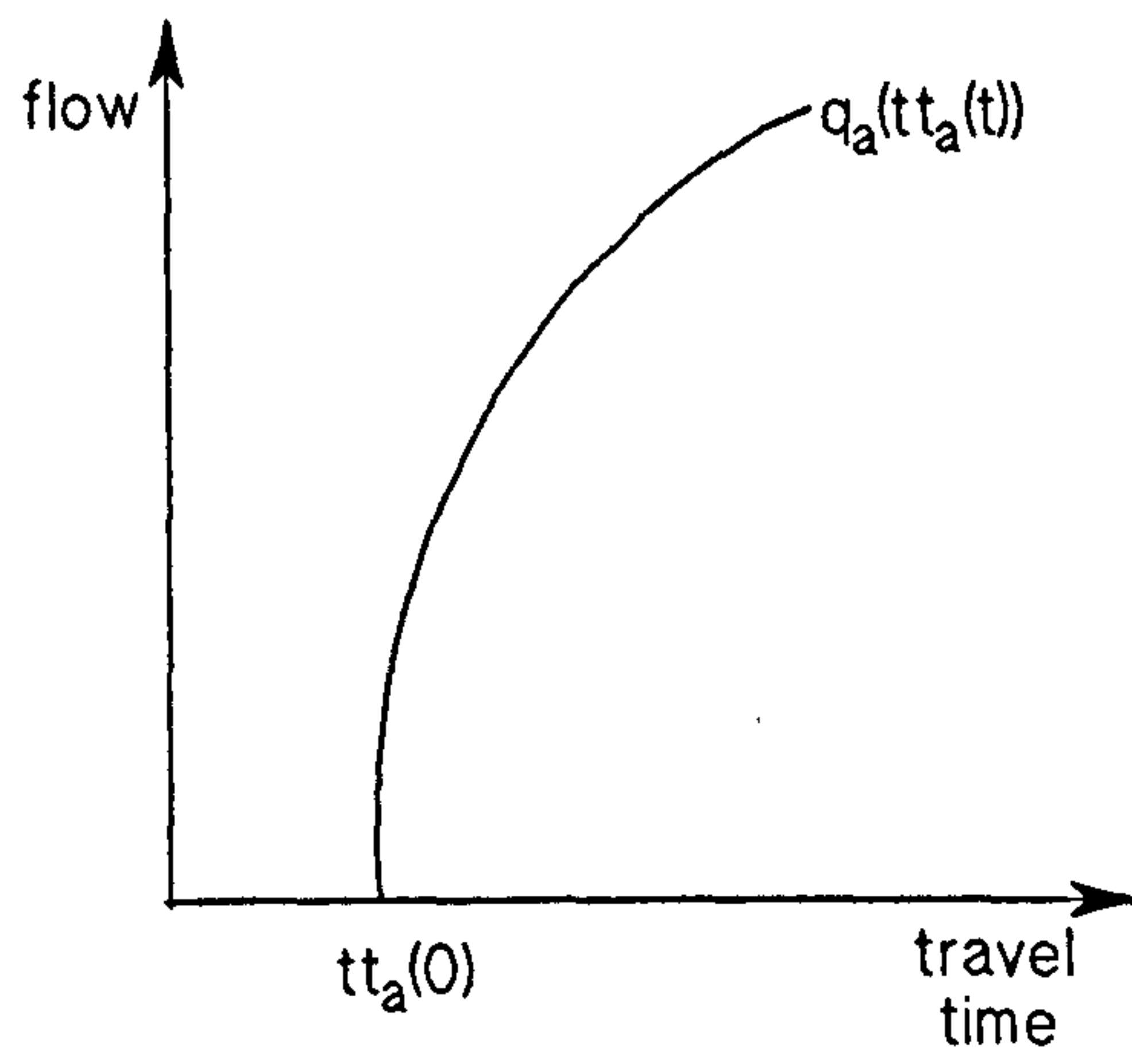
$$\frac{\partial}{\partial tt_b(t')} \left[ \sum_{rs} Q_{rs} S_{rs}[W_{rs}(t)] \right] = \\ = - \sum_{rs} Q_{rs} \sum_t \sum_k P_{k,rs}(t) \cdot \left[ \alpha \cdot \delta_{b,k}^{rs}(t',t) + \frac{\partial SD_{k,rs}(t)}{\partial tt_b(t')} \right] \quad (7.41)$$

Furthermore

$$\frac{\partial}{\partial tt_b(t')} \left[ \alpha \cdot \sum_t \sum_a \int_{tt_a(0)}^{tt_a(t)} q_a(u) du \right] = \alpha \cdot q_b(tt_b(t')) \quad (7.42)$$



(a)



(b)

Figure 7.1 : Inversion of the link performance function



Thus a typical term of the gradient  $\nabla_q z(tt)$  is given by

$$\frac{\partial z(tt)}{\partial tt_b(t')} = \alpha \cdot [q_b(t') - \sum_{rs} Q_{rs} \sum_t \sum_k P_{k,rs}(t) \delta_{b,k}^{rs}(t,t')] \quad (7.43)$$

Since the disutility (per unit time) due to travel time is a positive variable, i.e.  $\alpha > 0$ , then the gradient can vanish only when

$$q_b(t') = \sum_{rs} Q_{rs} \sum_t \sum_k P_{k,rs}(t) \delta_{b,k}^{rs}(t,t') \quad (7.44)$$

In other words when the time dependent link flows coincide with the DSUE flows. The gradient of  $z(tt)$  is expressed as:

$$\nabla_q z(tt) = \alpha \cdot [q - \sum_{rs} Q_{rs} P_{rs} (\Delta_{rs})^T] \quad (7.45)$$

Comparing equations (7.43) with (7.26), or (7.45) with (7.28), it should be noted that the gradients of  $z(q)$  and  $z(tt)$  have the same signs and vanish at the same points.

Using equation (7.43), the typical term,  $\partial z^2(tt)/\partial tt_b(t')\partial tt_a(t'')$ , of the Hessian of  $z(tt)$  is defined as :

$$\frac{\partial z^2(tt)}{\partial tt_b(t')\partial tt_a(t'')} = \alpha \cdot \left[ \frac{\partial q_b(t')}{\partial tt_a(t'')} - \sum_{rs} \sum_t \sum_k Q_{rs} \sum_i \sum_{t_i} \frac{\partial P_{k,rs}(t)}{\partial W_{i,rs}(t_i)} \cdot \frac{\partial W_{i,rs}(t_i)}{\partial TT_{i,rs}(t_i)} \delta_{a,k}^{rs}(t_i,t'') \cdot \delta_{b,k}^{rs}(t,t') \right] \quad (7.46)$$

Thus, the Hessian according to the above equation, can be expressed as the sum of two separate matrices as follows :

$$\nabla^2 z(tt) = \alpha \cdot \left[ \nabla_{tt} q + \sum_{rs} Q_{rs} \cdot \left[ (\Delta_{rs}) \cdot [(\nabla_{TT} W_{rs}) \cdot (-\nabla_W P_{rs})] \cdot (\Delta_{rs})^T \right] \right] \quad (7.47)$$

The first matrix in the right hand side of the above equation is  $\alpha \cdot \nabla_{tt} q$ . This is a  $\ell.n \times \ell.n$  diagonal matrix which includes terms of the form  $\alpha \cdot dq_a(t')/dtt_a(t')$  along its diagonal. The domain of the function  $q_a = q_a(tt_a)$  is defined as  $[tt_a(0), \infty)$ , and therefore  $q_a(tt_a)$  is an increasing function. Therefore, since  $\alpha > 0$  the terms  $\alpha \cdot dq_a(t')/dtt_a(t')$  are positive,  $\forall a, t'$ . Consequently the matrix  $\alpha \cdot \nabla_{tt} q$  is positive definite.

Following the arguments used earlier in this section, the second matrix in the sum that comprises the Hessian of  $z(tt)$ , includes a quadratic form of the positive semidefinite

matrix  $[(\nabla_{\mathbf{T}\mathbf{T}}\mathbf{W}_{\mathbf{r}\mathbf{s}}).(-\nabla_{\mathbf{W}}\mathbf{P}_{\mathbf{r}\mathbf{s}})]$ , and is therefore positive semidefinite. The Hessian  $\nabla^2z(\mathbf{t}\mathbf{t})$  is then the sum of a positive definite and a positive semidefinite matrix. This means that  $\nabla^2z(\mathbf{t}\mathbf{t})$  is positive definite and therefore,  $z(\mathbf{t}\mathbf{t})$  is strictly convex, and has a single minimum point.

The functions  $z(\mathbf{q})$  and  $z(\mathbf{t}\mathbf{t})$  are related by a monotonic transformation. In other words, each point of  $z(\mathbf{q})$  corresponds to one and only one point of  $z(\mathbf{t}\mathbf{t})$ . Furthermore the gradients of both functions always have the same sign and vanish at the same points. The gradient of  $z(\mathbf{t}\mathbf{t})$ , however, vanishes only once, at the minimum of  $z(\mathbf{t}\mathbf{t})$ . Thus  $z(\mathbf{q})$  must also have a unique minimum at the same point.

It should be noted that the uniqueness of the solution to the DSUE program (eq. (7.13)) is established with respect to the time dependent link flows, and not with respect to the time dependent path flows, as in the case of the Beckman formulation of the static deterministic user equilibrium, which was discussed in chapter 3.

### 7.5 Solution algorithm

In the previous sections of this chapter, it was shown that in order to solve the DSUE problem over a transportation network, it is sufficient to solve the minimisation problem given by equation (7.13)

In this section an algorithm for solving the DSUE equivalent program is presented. This algorithm is a *descent direction method* and is based on the approach proposed by Sheffi and Powell (1982) in their analysis of static stochastic equilibrium, which was briefly discussed in section 3.6.

Descent method algorithms generate a point  $\mathbf{q}^{n+1} = ( \dots, q_i(t_j)^{n+1}, \dots )$  from  $\mathbf{q}^n = ( \dots, q_i(t_j)^n, \dots )$ , so that  $z(\mathbf{q}^{n+1}) < z(\mathbf{q}^n)$ . This procedure is repeated until a predetermined convergence criterion is met, implying that the function  $z(\mathbf{q}^n)$  has reached its minimum at the point  $\mathbf{q}^n$ .

Thus, as was mentioned in section 3.5, where a solution algorithm (Frank and Wolfe) for the deterministic static assignment was described, the basic algorithmic step in a minimisation procedure can be expressed as

$$\mathbf{q}^{n+1} = \mathbf{q}^n + \lambda^n \cdot \mathbf{d}^n \quad (7.48)$$

where :

- $\mathbf{q}^n$  is the vector of decision variables, i.e. the time dependent link flows in the DSUE equivalent minimisation problem, at the  $n$ th iteration,
- $\mathbf{d}^n$  is a descent direction vector calculated at  $\mathbf{q}^n$ ,

$\lambda^n$  is a nonnegative scalar representing the move size, i.e., it determines how far along  $d^n$  the next point  $q^{n+1}$  will be.

For a function  $z(q_1, q_2, \dots, q_i)$ , a descent direction at some point  $q' = (q_1', q_2', \dots, q_i')$ , can be given by the opposite to the direction of the gradient at that point. Thus for the minimisation program (7.13), a descent direction can be expressed as the opposite of the gradient, given by equation (7.28), as follows :

$$d^n = \alpha \cdot \left[ \sum_{rs} Q_{rs} \cdot P_{rs}^n \cdot (\Delta_{rs}^n)^T - q^n \right] \cdot \nabla_{q} z \quad (7.49)$$

A simpler descent direction can be defined by omitting  $\alpha$  and the Jacobian  $\nabla_{q} z$  from the above product. Thus  $d^n$  can be expressed as :

$$d^n = \sum_{rs} Q_{rs} \cdot P_{rs}^n \cdot (\Delta_{rs}^n)^T - q^n \quad (7.50)$$

$d^n$  as expressed by equation (7.50) is a descent direction vector, since the product  $\nabla z(q^n) \cdot d^{nT}$  is always negative :

$$\nabla z(q^n) \cdot d^{nT} = -\alpha \cdot \left[ \sum_{rs} Q_{rs} \cdot P_{rs}^n \cdot (\Delta_{rs}^n)^T - q^n \right] \cdot \nabla_{q} z \cdot \left[ \sum_{rs} Q_{rs} \cdot P_{rs}^n \cdot (\Delta_{rs}^n)^T - q^n \right]^T \leq 0$$

Furthermore because of the uniqueness of the minimum,  $\nabla z(q^n)^T d^n$  is strictly negative except where  $\nabla z(q^n)$  vanishes.

A typical element of the descent vector given by (7.50) is expressed as:

$$d_a(t')^n = \sum_{rs} Q_{rs} \sum_t \sum_k P_{k,rs}(t)^n \delta_{a,k}^{rs}(t,t')^n - q_a(t)^n \quad (7.51)$$

where  $P_{k,rs}(t)^n$  is the probability of departing at time  $t$  and selecting route  $k$  as determined from the utility distributions at the  $n$ th iteration. In other words :

$$P_{k,rs}(t)^n = \Pr[ U_{k,rs}(t)^n > U_{m,rs}(t')^n \quad \forall (k,t) \neq (m,t'), \quad k,m \in K_{rs}(t)^n ] \quad (7.52a)$$

or equivalently

$$P_{k,rs}(t)^n = \Pr[ W_{k,rs}(t)^n < W_{m,rs}(t')^n \quad \forall (m,t') \neq (k,t), \quad k,m \in K_{rs}(t)^n ] \quad (7.52b)$$

and  $P_{rs}^n = ( \dots, P_{k,rs}(t)^n, \dots )$  is the vector of these probabilities.

The components of the descent direction vector, as shown in equation (7.51), are defined as the difference between two quantities. Therefore, each component can be written in a form similar to the one used in the Frank and Wolfe algorithm, (i.e. as

described in the paragraph following eq. 3.19) . Thus, let  $y_a(t')^n$  be the auxiliary variable associated with link a and time t' :

$$y_a(t')^n = \sum_{rs} Q_{rs} \sum_t \sum_k P_{k,rs}(t)^n \cdot \delta_{a,k}^{rs}(t,t')^n \quad \forall a, t' \quad (7.53)$$

Each component of the descent direction vector can then be written as :

$$d_a(t)^n = y_a(t)^n - q_a(t)^n \quad (7.54)$$

or, in vector form, as

$$d(t)^n = y(t)^n - q(t)^n \quad (7.55)$$

The set  $\{y_a(t')^n\}$  of the auxiliary variables defined by equation (7.53) is a function of  $P_{rs}^n$ , which in turn, as shown in equation (7.52), depends on the pattern of the time dependent distribution of the total disutility (utility) of travel,  $W_{rs}^n$ . Therefore  $\{y_a(t')^n\}$  can be obtained by a dynamic stochastic network loading which is based on  $W_{rs}^n$ .

Having presented the methodology which can be used to calculate the descent direction vector  $d^n$ , the next step in the optimisation procedure is to define the move size  $\lambda^n$  which will determine the distance along the descent direction from the current solution point  $q^n$  to the next solution point  $q^{n+1}$ . The move size is typically chosen so that the objective function is minimised along the descent direction. In other words  $\lambda^n$  is defined from the following one-dimensional minimisation problem

$$\min_{\lambda^n} z[ q^n + \lambda^n \cdot (y^n - q^n) ] \quad (7.56)$$

The solution of the minimisation program (7.13) can then be obtained by following an iterative process of finding the descent direction and minimising the objective function along this direction.

Looking at equation (7.13), it should be noted, that the form of the objective function is general so that to include any stochastic choice model. More specifically, the component of the first term on the right hand side of equation (7.13), expressing the expected total disutility of travel, can be defined not by only using a logit formulation, (as was used in the previous chapter), but any stochastic multiple choice model; for example it might be defined by using a probit formulation. However in the case, that the probability of selecting a certain alternative is not calculated by an analytical expression, but simulation has to be used, these probabilities and consequently the pattern of the time dependent link flows can only be estimated, rather than computed accurately. Thus, the vector of the auxiliary link flows  $y^n$ , (eq. (7.53)) is a random

variable, implying that the descent direction (eq. (7.55)) is also random. Therefore, although the vector  $\mathbf{d}$  is on average a descent vector, the vector  $\mathbf{d}^n$  computed at a particular iteration,  $n$ , may not point at a descent direction. This may cause problems in applying any descent direction method.

Thus a solution algorithm which will converge to the minimum even if the random vector  $\mathbf{d}$  is not a descent direction vector at each iteration would be very useful, since it would enable the use of the DSUE equivalent minimisation formulation, with any stochastic multiple choice model. Such an algorithm is the method of successive averages, which is analysed below.

### 7.5.1 The Method of Successive Averages

The method of successive averages is a descent direction method, in which the move size  $\lambda$  is not defined from the minimisation program (7.56); instead the sequence  $\lambda^1, \lambda^2, \dots$  is determined a priori.

This method is based on the approach suggested by Blum (1954), who developed a solution algorithm for stochastic optimisation problems in which the objective function has continuous first and second partial derivatives, and its gradient vanishes only once. The algorithmic step is given by equation (7.48), and the sequence of move sizes have to satisfy the following conditions :

$$\sum_{n=1}^{\infty} \lambda^n = \infty \quad (7.57a)$$

$$\sum_{n=1}^{\infty} \lambda^{n^2} < \infty \quad (7.57b)$$

Blum (1954) has proved that if the above conditions are satisfied, then the sequence of the solutions  $\mathbf{q}^{n+1}, \mathbf{q}^n, \dots$  will converge to the vector which provide the minimum of the objective function, even if the descent direction vector is a random vector. This algorithm will converge to the optimum solution, subject to the condition that on average the descent direction vector strictly points at a descent direction.

The harmonic sequence  $1, 1/2, 1/3, 1/4$  etc is a simple move size sequence that satisfies both conditions (7.57). This is the sequence that is used in the method of successive averages, and can be expressed as:

$$\lambda^n = 1/n \quad (7.58)$$

Using the above formula for the move size, the basic algorithmic step, as it is given by equation (7.48), can be expressed as :

$$q^{n+1} = q^n + (1/n).d^n \quad (7.59)$$

Using the form of the descent direction vector as it is expressed by equation (7.55), the nth algorithmic step can be defined by the following equation :

$$q^{n+1} = q^n + (1/n).(y^n - q^n) \quad (7.60)$$

The above equation implies that the solution at each iteration is the average of the variables y in all the preceding iterations, since

$$\begin{aligned} q^{n+1} &= \frac{n-1}{n} q^n + \frac{1}{n} y^n \\ &= \frac{n-1}{n} \cdot \frac{n-2}{n-1} q^{n-1} + \frac{n-1}{n} \cdot \frac{1}{n-1} y^{n-1} + \frac{1}{n} y^n \\ &= \frac{n-2}{n} q^{n-1} + \frac{1}{n} (y^{n-1} + y^n) \\ &= \frac{1}{n} \sum_{k=1}^n y^k \end{aligned} \quad (7.61)$$

In order the MSA to be an accurate minimisation algorithm, it has :

- to converge to the minimum, and also
  - to dissipate the errors resulting from the use of a random search direction procedure.
- The proof of the above conditions is described below, following the arguments used by Sheffi (1985) for the case of static stochastic assignment.

Equation (7.57a) guarantees that the algorithm will not stop before the minimum is reached. This happens because the algorithm has the ability, at every iteration m to move from the current point  $q^m$  to any other feasible point and particularly towards the minimum. This is because the condition (7.57a) implies that

$$\sum_{n=m}^{\infty} \lambda^n = \infty \quad \text{for any positive integer } m$$

This might not happen if the sequence were, for example, expressed as  $\lambda^n = 1/n^2$ , in which case the algorithm might terminate before the minimum is reached.

Furthermore, if the search direction is random then the flow pattern at the nth iteration, denoted by  $q^n = (. . . , q_a(t)^n, . . .)$  is a random variable. The condition (7.57b)

guarantees that the variance of the random variable  $\{ q_a(t) \}$  will become smaller and smaller as the number of iterations increases:

$$\text{var}(q_a(t)^{n+1}) = (1/n^2) \sum_{m=1}^n \text{var}(Y_a^m)$$

where  $\text{var}(Y_a^m)$  is the variance of the random auxiliary flow on link  $a$  at time  $t$ .

The right hand side of the above equation approaches zero as  $n$  increases, implying that the variance approaches zero as the algorithm progresses. This happens because the variance of  $Y_a^n$  is bounded from above by some value  $x < \infty$ , since the auxiliary flow  $Y_a^n$ , given by equation (7.53), must be positive and as well as less than the sum of all the O-D demand.

Thus as was demonstrated above, a sufficient condition for the algorithm to converge, is the search direction to be a descent vector only on the average.

The steps of the MSA algorithm can then be defined as follows :

**Step 0: Initialisation.**

Set  $v_a(t)^0 = v_a \quad \forall a \in L, t \in [T_0, T_0+T]$

Calculate  $tt_a(t)^0 = tt_a(v_a(t)^0)$ .

Perform a dynamic stochastic network loading based on the set of the initial free flow travel times  $\{tt_a(t)^0\}$ .

This generates a set of time dependent link flows  $\{q_a(t)^1\}$ .

Set  $n=1$

**Step 1: Update.**

Set  $v_a(t)^n = v_a(q_a(t)^n) \quad \forall a \in L, t \in [T_0, T_0+T]$ .

Calculate  $tt_a(t)^n = tt_a(v_a(t)^n)$

**Step 2: Direction Finding.**

Perform a dynamic stochastic network loading procedure based on the current set of the time dependent link travel times  $\{tt_a(t)^n\}$ .

This yields the auxiliary time dependent link flow pattern  $\{y_a(t)^n\}$

**Step 3: Move.**

Find the new time dependent flow pattern by setting

$$q_a(t)^{n+1} = q_a(t)^n + (1/n) \cdot [y_a(t)^n - q_a(t)^n] \quad \forall a \in L, t \in [T_0, T_0+T]$$

**Step 4: Convergence Criterion.**

If the convergence criterion is met, stop; the DSUE flows are the time dependent flow patterns calculated in step 3.

Otherwise set  $n=n+1$  and go to step 1.

As Sheffi (1985) states, it can be argued that the sequence  $\{\lambda^n\}$  used in the MSA, forces the sequence  $\{q^n\}$  to converge. Thus criteria which are based on the relative change of link flows may not be appropriate for use in the MSA. A convergence criterion can then be based on the rate of the reduction in the objective function, or alternatively the algorithm may stop after a predetermined number of iterations. A convergence measure based on flow similarity can be obtained if this measure is based on the flow in the last several iterations. Such a convergence criterion can be expressed as:

$$\frac{\sqrt{\sum_t \sum_a (q_a'(t)^{n+1} - q_a'(t)^n)^2}}{\sum_t \sum_a q_a'(t)^n} < \xi$$

where  $q_a'(t)^n$  expresses the flow average over the last  $m$  iterations and is defined as :

$$q_a'(t)^n = \frac{1}{m} \sum_{x=0}^{m-1} q_a(t)^{n-x}$$

## 7.6 Summary

In this chapter a different approach for solving the dynamic assignment problem was proposed. Based on the approach used in the static stochastic assignment, a framework for formulating and solving the DSUE problem as an equivalent optimisation program was presented. The derivation is however only true under certain restrictive assumptions. Thus the analysis presented should be regarded as a framework for formulating the dynamic assignment as an equivalent program, rather than as a strict formulation. However, in terms of empirical evidence, the results derived from the proposed algorithm and the simulation procedure that defines the equilibrium traffic patterns by adjusting the demand patterns from day to day (chapter 6) are shown to be identical (section 8.7).

The proposed algorithm involves only the calculation of the time dependent link flow patterns, implying that the time dependent path flows,  $Q_{k,rs}(t)$  do not need to be defined. Therefore, the major advantage of the approach developed in this chapter is that the proposed algorithm, in conjunction with a DSNL mechanism which does not require path enumeration, would substantially increase the computational efficiency of the algorithm in terms of memory requirements. A stochastic network loading mechanism which obviates path enumeration was developed by Dial (section 3.6.1.) for the static assignment problem. However till now, no such mechanism has been



developed for the dynamic assignment problem; the DSNL mechanism which can be derived from the model developed in the previous chapter can handle general networks, but it could be criticised as not a computationally efficient approach. The proposed framework will therefore enable the analysis of large networks, if efficient DSNL mechanisms are developed.

# 8 experimental analysis of the DSUE model

## **Objective**

The aim of this chapter is to illustrate the outputs of the DSUE assignment model. The chapter demonstrates the features of the model outputs and investigates the factors that generate these features.

### **8.1 Introduction**

In the last two chapters two approaches for solving the DSUE assignment problem have been developed. The simulation model is coded in FORTRAN 77 and implemented on a VAX 8850. Changes in the parameters and the specifications of the initial conditions can be easily made. The model produces detailed information on the last iteration of the model, and historical information for every iteration of the simulation, or less frequently at iterations specified by the user.

This chapter is dedicated to the analysis of the developed model outputs in order to demonstrate its capability to replicate the time varying traffic patterns during the morning peak, and to test its sensitivity to a number of variables it includes. The analysis is based on the results of simulation based experiments involving travellers driving from home to work during the morning peak period. The presentation of the results relating to the overall system's performance is structured into six sections.

In section 8.2 the commuting context used for the simulation experiments is introduced. The different outputs of the DSUE model are presented to demonstrate the ability of the model to describe the time varying traffic patterns during the morning peak period. The time dependent departure rate and O-D travel time distributions are presented to illustrate the peaking characteristics of travel demand. The factors that determine the

time and space distribution of traffic patterns are analysed and the evolution of congestion within the network links is described.

Section 8.3 demonstrates the sensitivity of the model outputs to the level of the work start time flexibility. Inflexible working hours result in high peaks in the departure rate distributions, since travellers want to arrive at their destinations at the same time. The high concentration of traffic within a short time interval results in long delays and higher levels of schedule delay. On the other hand, flexible working hours cause a spreading of the demand over a longer time interval and therefore reduce delays.

In section 8.4 the sensitivity of the model to the values of the parameters reflecting the variability of preferences with respect to route and departure time choice is analysed. The model outputs demonstrate that increased variability of preferences results in a wider spread of the demand in time and space and therefore to lower travel times but longer schedule delays. Lower levels of variability of preferences imply that drivers are more concentrated to optimum routes and departure times.

The effects that the level of demand have on the traffic patterns during the morning peak are analysed in section 8.5. The departure rate distributions under the increased level of demand are more highly peaked and shifted towards earlier departures reflecting that the travellers respond to the higher levels of congestion by shifting towards earlier departure times.

In section 8.6 the DSUE model that uses the travel time model which is based on queueing theory is applied. The commuting context used in this experiment is briefly described. The model demonstrates its ability to represent the build up and dissipation of queues created at bottlenecks in urban networks.

Finally in section 8.7 the results from the application of the demand adjustment mechanism based approach and the MSA approach are compared. The outputs were found to be identical.

## **8.2 The base case experiments using the traffic flow model**

The network used for the simulation experiments conducted using the traffic flow based travel time model has the form of the Sioux Falls, South Dakota, network. This network has already been used by other researchers (LeBlanc (1975), Poorzahedy (1980), Haghani and Daskin (1983) and others) in the analysis of network design

problems and also previously by the author (1989, 1990a) to demonstrate the outcomes of the DSUE model. The network, consists of 23 nodes and 72 links and is depicted in figure 8.1. The speed-density relationship used in the base case is the one developed from the BPR travel time model (eq. 6.22) and the values of the parameters  $a$  and  $b$  were considered the same for all the links and equal to 0.15 and 4.0 respectively. The total demand using the network is assumed to be constant and the travellers have the choice among alternative departure times and routes. Network users are assumed to travel between 12 O/D pairs shown in table 8.1. The values of the parameters  $\alpha$ ,  $\beta$  and  $\gamma$  used in the utility model are derived from Small (1982). Furthermore, the values of the scale parameters that reflect the variability of preferences among travellers are  $\mu_r = \mu_t = 1$ . The desired arrival time period was considered to be the same for all destinations. The base case desired arrival time period is defined as [8:45 am to 9:15 am] implying that the work starting time flexibility is  $D_{rs} = 15$  minutes and  $t_{rs} = 9:00$ . The period of the day analysed (defined by the variables  $T_0$ , and  $T_0+T$ ) is [7:00 am to 10:00 am]. The time increment  $h$  used in the simulation model is 1 min, and the reviewing rates  $F_1$  and  $F_2$  are 0.15 and 0.35 respectively. The latter values imply that every day 15% of the travellers review their previous trip decisions, and 35% of them will change both or at least one of their dimensions of choice. The value of the tolerance factor  $\xi$ , defining the convergence of the algorithm, has taken the value 0.2, in other words convergence of the system towards an equilibrium state is assumed to occur when the maximum relative deviations of the departure rate from one day to the next is less than 3%, ( $=\xi.F_1$  from eq. 6.72).

The base case simulation experiment begun with no-congestion pseudoequilibrium departure rate distributions; the stationary state of the system for the base case simulation experiment was reached after 39 iterations when the convergence criterion  $RR < \xi=0.2$  was met. To demonstrate the convergence of the system towards the equilibrium state, the dynamic evolution of the departure rate distribution towards the stationary state for O-D pair 4-20 is depicted in figure 8.2, and the maximum relative variation of the departure rate from one iteration to the next is plotted against of the number of iterations performed in figure 8.3 . For the first 10 iterations the convergence pattern exhibits an oscillatory behaviour. After iteration 13 this pattern is rather smooth and "well behaved". The rate of decrease of the maximum relative variation  $RR$  is kept at higher levels for the band of iterations 13 to 19, and then  $RR$  decreases at an almost constant rate.

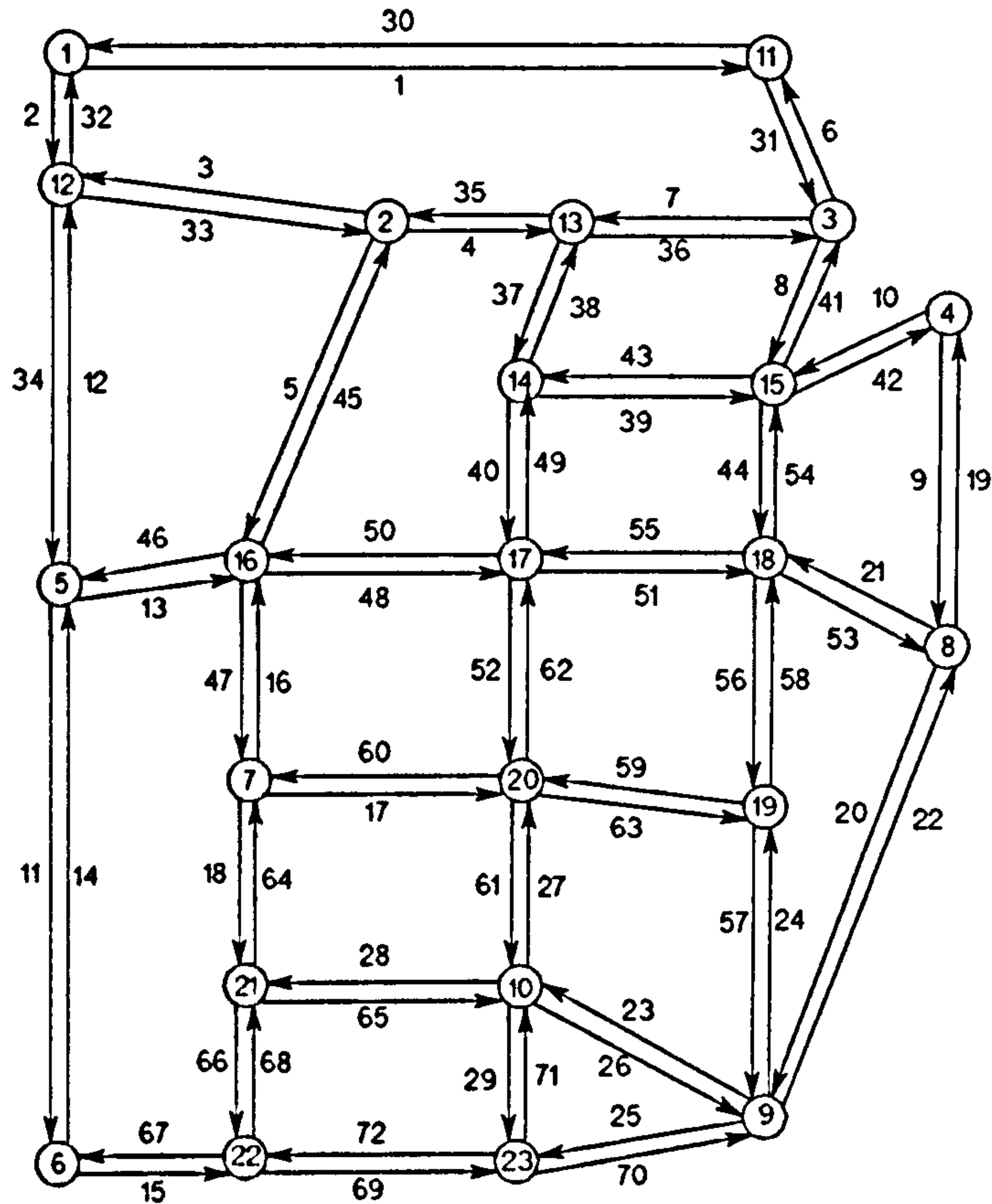


Figure 8.1 : The Sioux Falls network

(a description of the network data is given in the appendix A2)

Table 8.1: Origin - Destination pairs and demand.

O-D PAIR	DEMAND
1 - 17	2900.
2 - 19	2800.
3 - 20	2600.
4 - 20	2800.
5 - 19	2600.
6 - 17	2300.
7 - 15	2200.
8 - 13	2800.
9 - 14	2700.
10 - 15	2800.
11 - 20	2600.
12 - 18	2700.

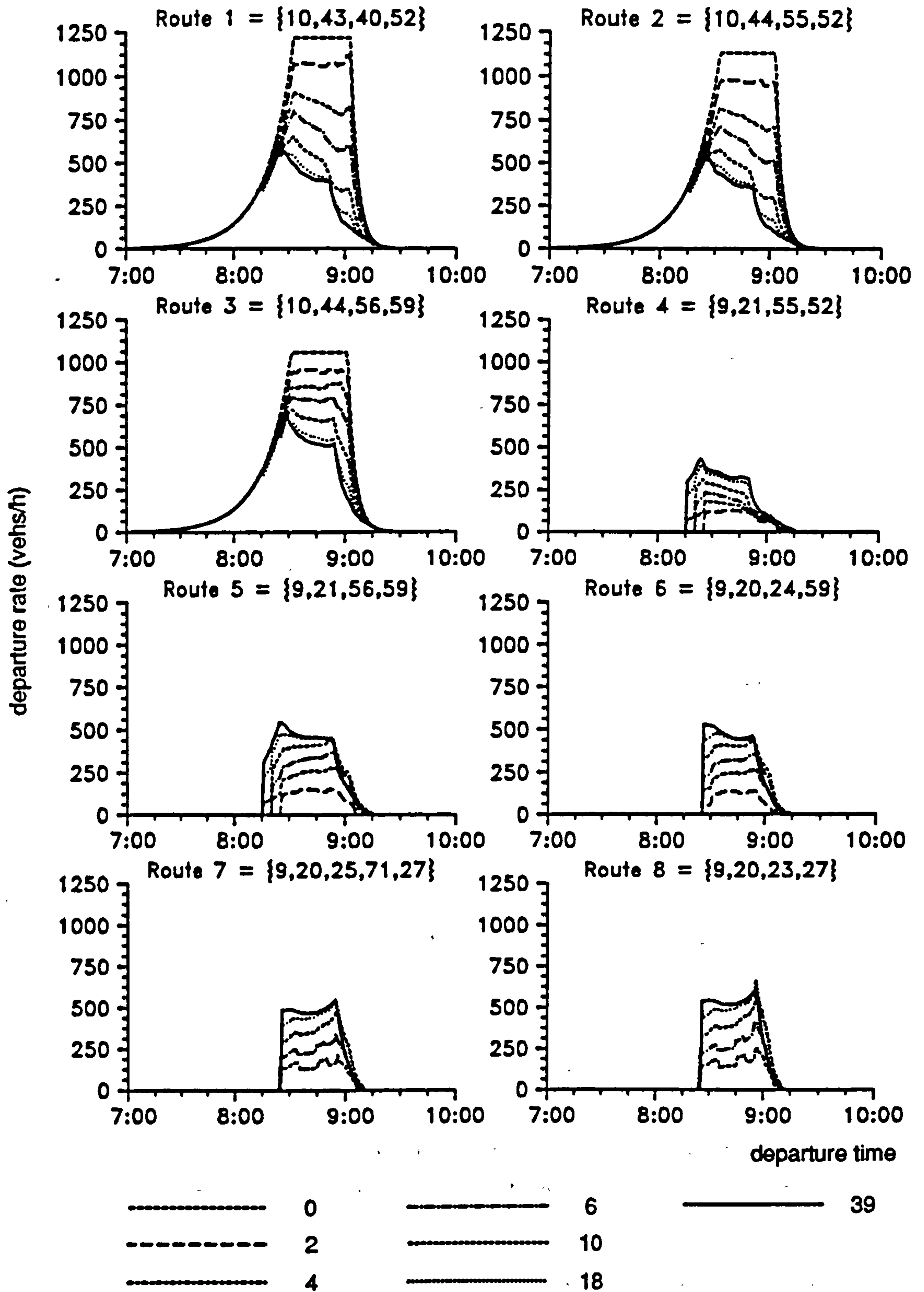


Figure 8.2: Evolution of departure rate distributions towards the equilibrium state for O-D pair 4-20.

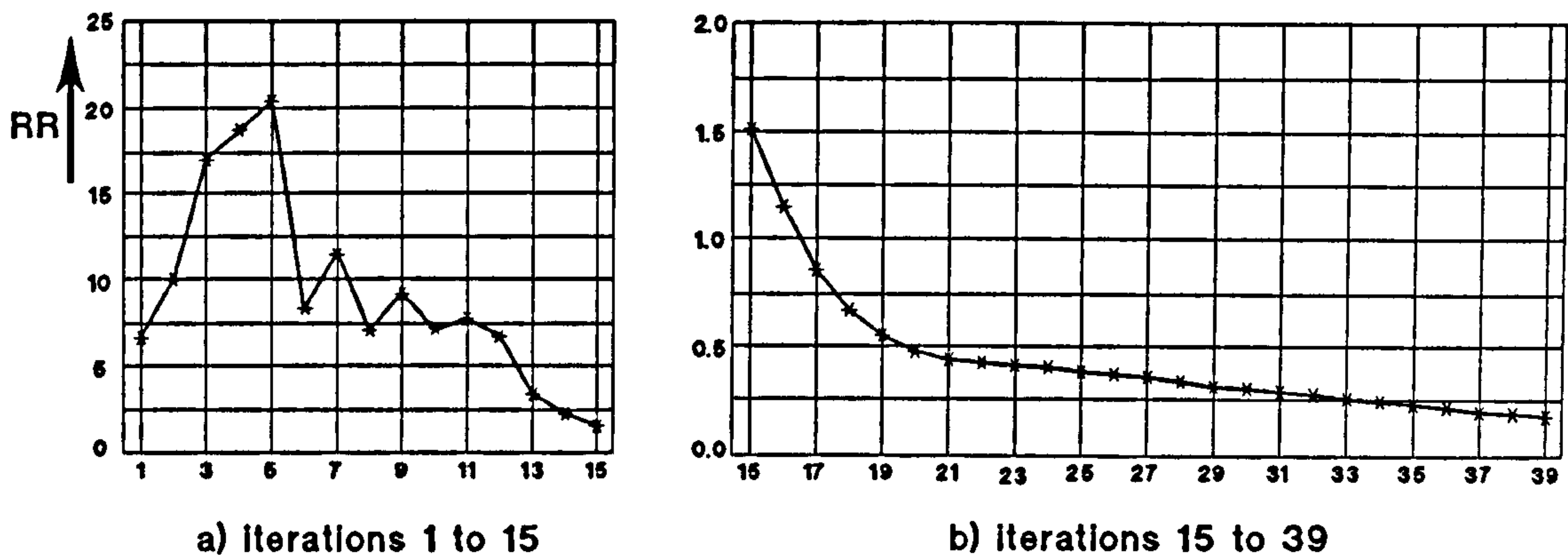


Figure 8.3 : Convergence pattern for the base case.

To illustrate the model outputs for the base case experiment, the equilibrium distributions of departure rates and travel times for the O-D pairs 2-19, 4-20 and 7-15 are depicted in figures 8.4 8.5 and 8.6. The figures demonstrate that travel times are not constant during the peak as is usually assumed in static traffic assignment procedures. As an example, the O-D pair 2-19 is examined below in more detail, by referring to figure 8.7.

At about 8:00 free flow conditions still exist for all the available routes connecting the origin (2) to destination (19). After that time congestion builds up rapidly in all the routes. An illustration of the level of congestion associated with each route is demonstrated in figure 8.8 which presents the free flow travel times and the difference between free flow and peak travel times on the routes connecting the O-D pair 2-19.

Routes 6 and 7, are associated with the highest free flow travel times, and become severely congested during the peak period, mainly due to the low capacity to demand ratio in links they include. For instance this ratio for links 48, 52 and 56 are 2800/4985, 2500/5059 and 2700/3843, respectively. As shown in figure 8.7, travel time along routes 6 and 7 are the highest compared to the ones along alternative routes, resulting in the lowest peaked departure rate distributions on routes 6 and 7 as travellers shift to alternative routes.

Route 4 is the most congested route. The peak travel time is 89% longer than the free flow travel time mainly due to the high level of congestion in links 40, 52 and 63 link. Figure 8.8 illustrates the difference between free flow and peak travel time which is about 12 mins, and is the highest among the alternative routes. However, route 4 is one of the two routes associated with the lowest free flow travel time, and therefore



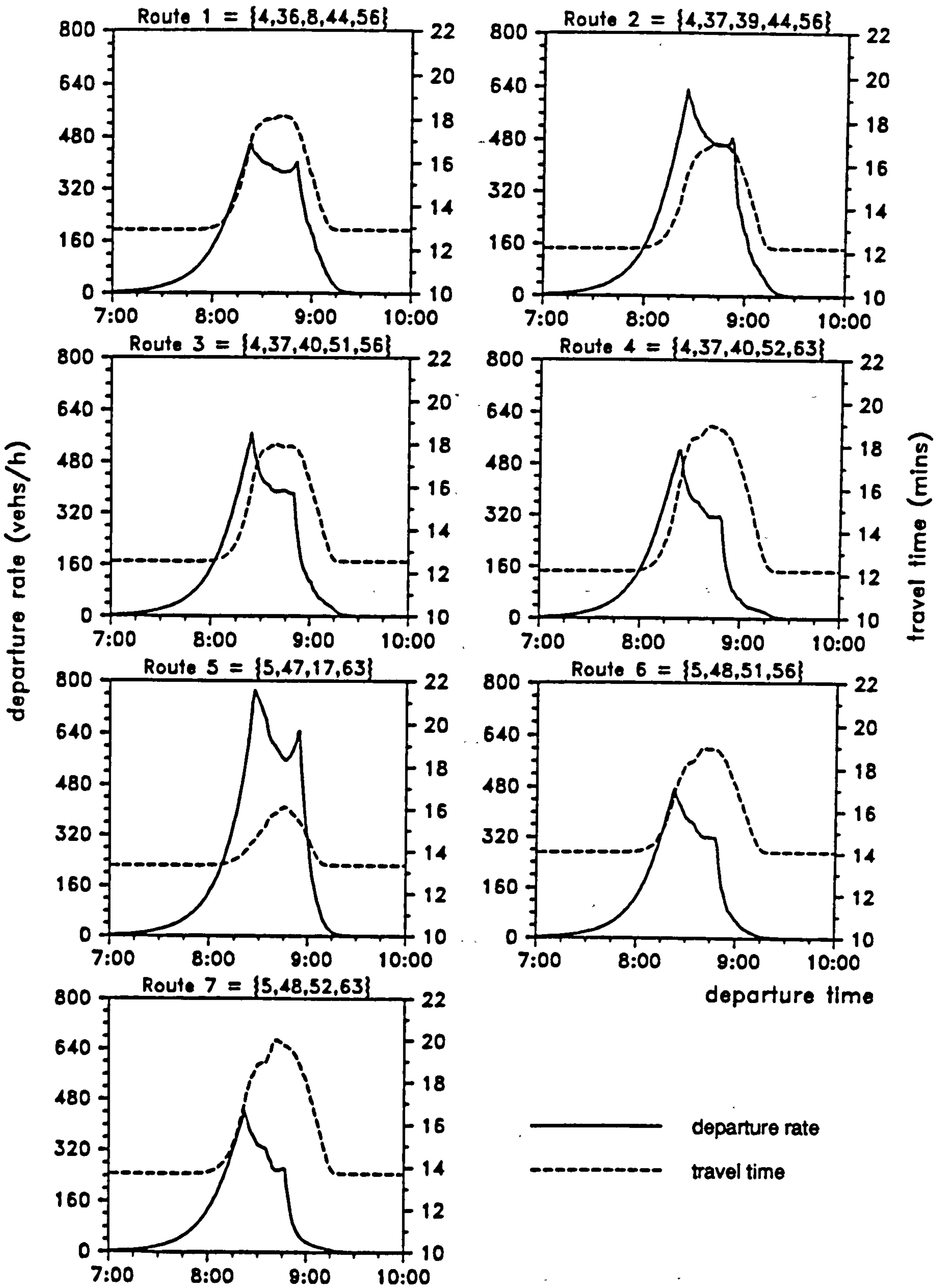


Figure 8.4: Departure rate and travel time distributions for O-D pair 2-19.

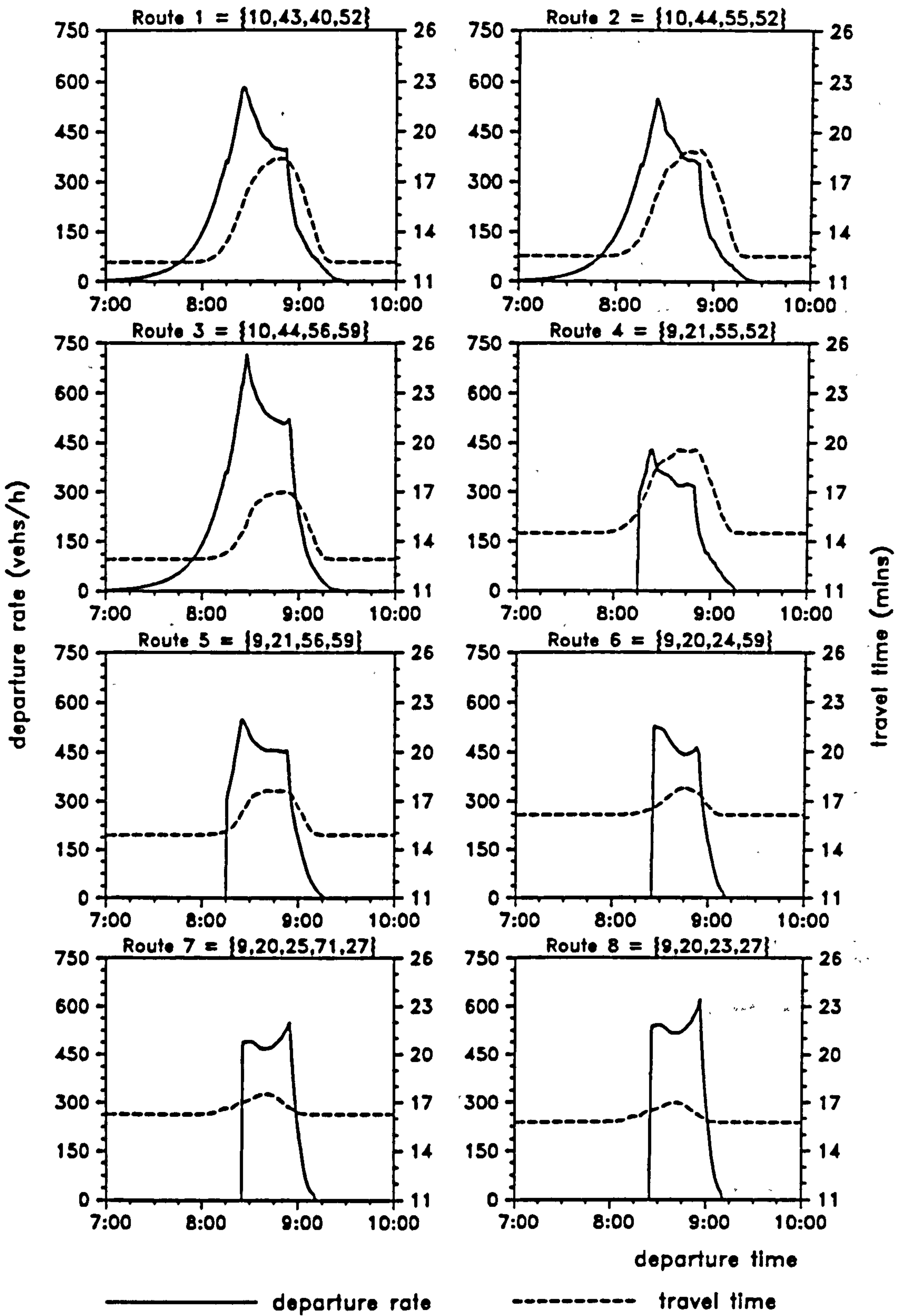


Figure 8.5: Departure rate and travel time distributions for O-D pair 4-20.

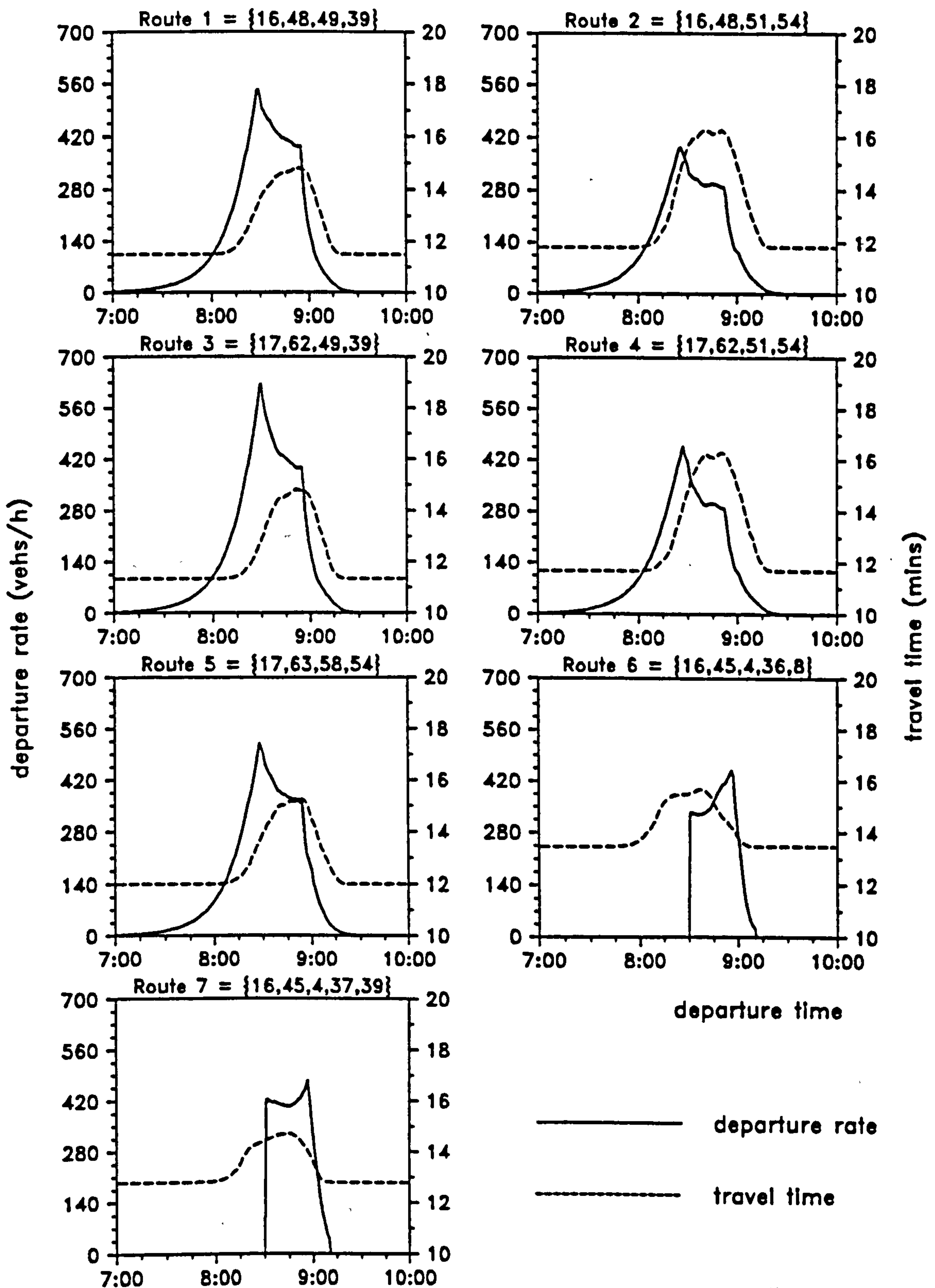


Figure 8.6: Departure rate and travel time distributions for O-D pair 7-15.

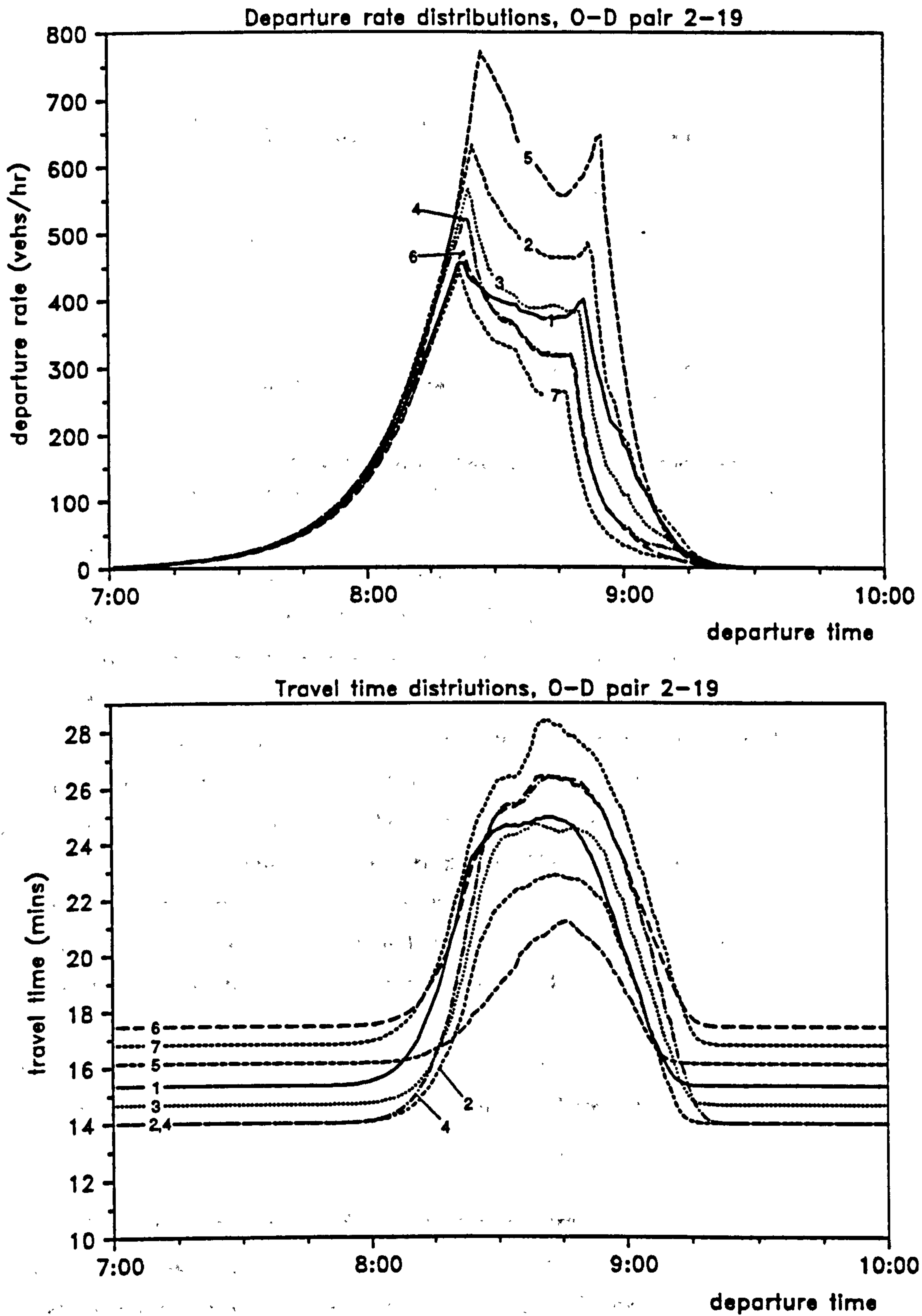


Figure 8.7: Departure rate and travel time distributions for O-D pair 2-19.

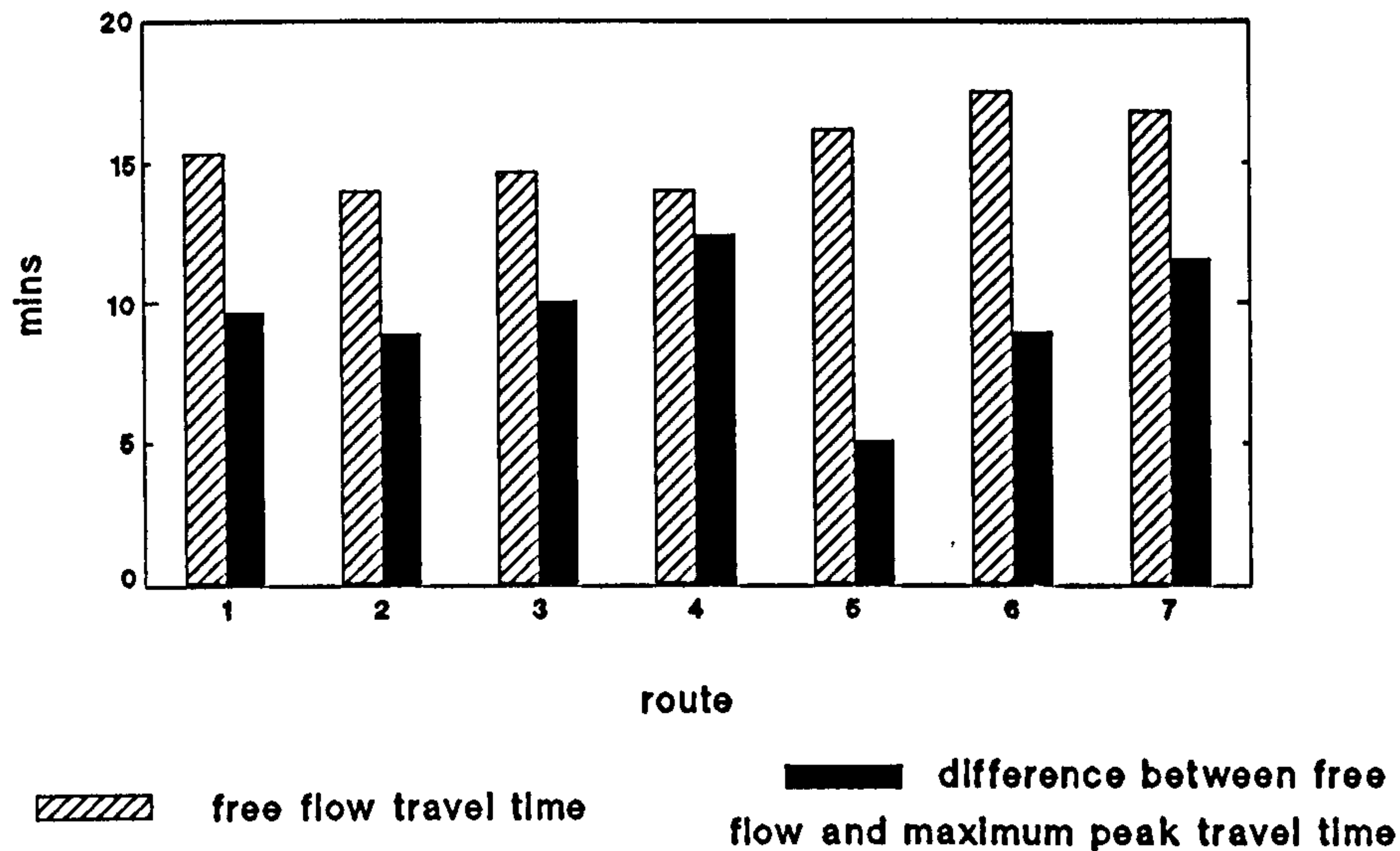


Figure 8.8 : Free flow travel times and difference between free flow and maximum peak travel times on routes connecting the O-D pair 2-19

although is the most congested, it is not associated with the highest travel times. For departures after 8:20 the travel time distribution along route 4 is very similar to route 6 and their resulting departure rate distributions are almost identical. Route 5 is the less congested route since the difference between peak and free flow travel time is just above 5 mins. Thus although it is not associated with the lowest free flow travel time, it offers the best level of service (defined in terms of travel time) during the peak period. It therefore attracts the highest proportion of the demand, and is associated with the highest departure rates and highest peaks. The difference between peak and free flow travel time in route 2 is about 9 mins. Although this difference seems to be high when compared to the one associated with route 5, route 2 provides a good level of service and is characterised by relatively high peaks in the departure rate distribution, since it is one of the two lowest free flow travel time routes. Routes 1 and 3 are characterised by relatively high levels of congestion and low free flow travel times. The distribution of departure rates and travel time on these routes follow a pattern that lies in between those experienced on routes 4,6,7 and 2,5.

The equilibrium distributions of departure rates, travel times, disutilities and schedule delays for routes 1 and 5 connecting the O-D pair 2-19 are presented in figure 8.9. Travel time distributions have a smooth shape, while departure rate distributions are continuous but do not have continuous derivatives due to the fact that the utility function has different left and right derivatives at  $t_{k,rs}^1$  and  $t_{k,rs}^2$ . The resulting stationary distributions have similar shape to the ones already derived by Ben-Akiva *et*

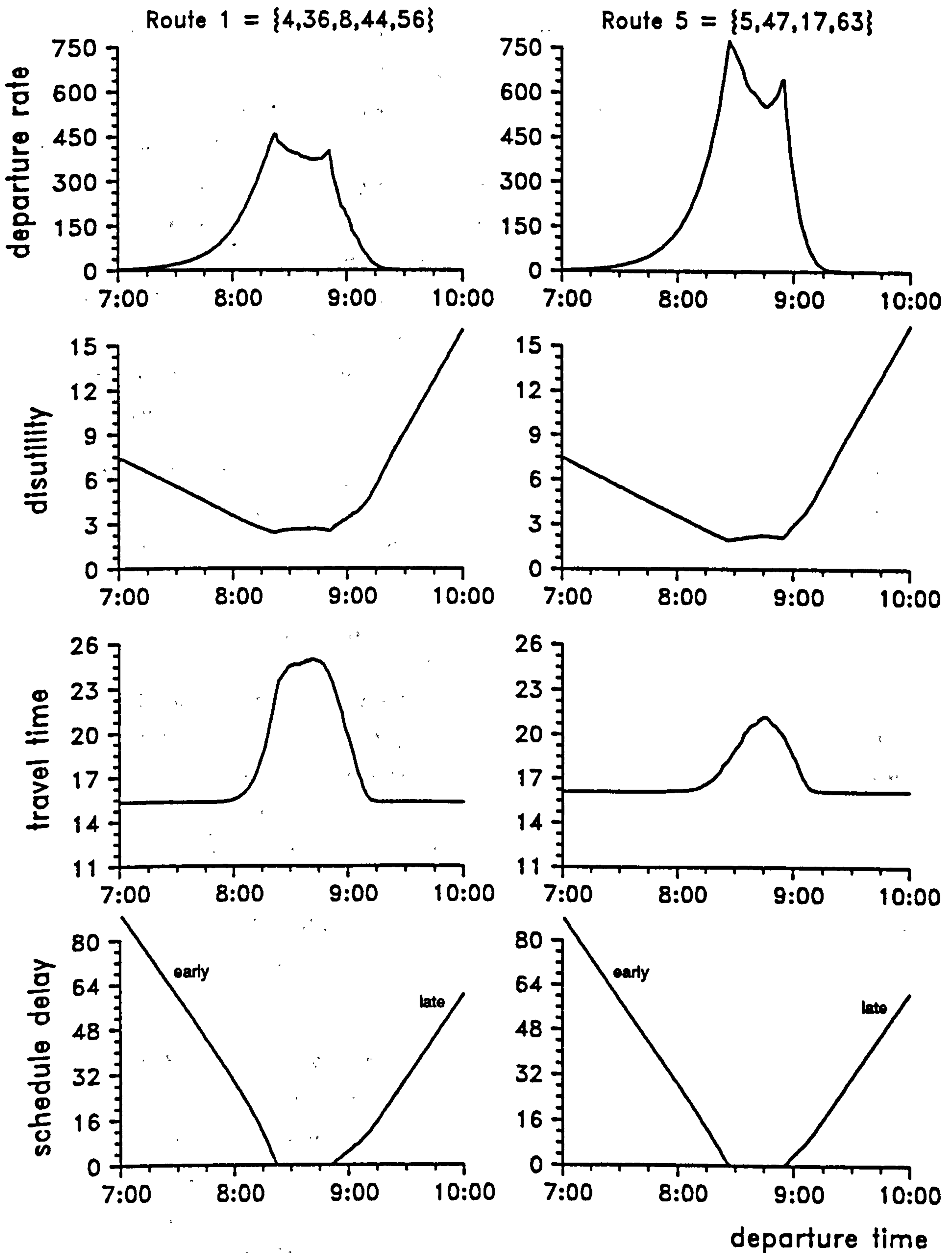


Figure 8.9: Departure rate, disutility, travel time and schedule delay disutilities for routes 1 and 5 connecting the O-D pair 2-19.

*al.* (1984,1986,1986a) and de Palma *et al.* (1983,1984). They are not consistent with Wardrop's first principle adjusted to time dependent trip characteristics, since in contrast to Wardrop's deterministic approach, a stochastic one was adopted in this thesis. For both routes the periods of departure times for on-time arrivals [  $t_{1,2-19}^1, t_{1,2-19}^2$  ] = [ 8:23, 8:51 ] and [  $t_{5,2-19}^1, t_{5,2-19}^2$  ] = [ 8:27, 8:55 ] occur within the congestion period, i.e. when travel times are higher than free flow travel times.

For departures earlier than 8:00, i.e. for free flow travel conditions and early arrivals, the schedule delay is a linear decreasing function of the departure time, while for departures later than 9:15, i.e., for free flow travel conditions and late arrivals, the schedule delay is a linear increasing function of the departure time. Within the congestion period, travel time and consequently schedule delay distributions do not follow linear patterns.

During the interval [  $T_0, t_{k,rs}^1$  ] the total disutility of travel is continuously decreasing, implying that the rate of increase of the disutility due to travel time is lower than the rate of decrease of the disutility due to schedule delay; the departure rate distribution is increasing exponentially and at time  $t_{k,rs}^1$  attains its maximum.

During the interval [  $t_{k,rs}^1, t_{k,rs}^2$  ] the distribution of the total disutility of travel follows the pattern of the travel time distribution. This because the total disutility is attributable only to travel time since departures during that period result in on-time arrivals. After the departure rate achieves its maximum at time  $t_{k,rs}^1$ , it starts to continuously decrease up to the point in time that travel time attains its maximum. Then the departure rate distributions follow an increasing pattern and exhibit a second but lower peak at time  $t_{k,rs}^2$ . After time  $t_{k,rs}^2$ , the disutility of travel starts increasing since the rate of decrease of the disutility due to travel time is lower than the rate of increase of the disutility due to schedule delay. Thus during the interval [  $t_{k,rs}^2, T_0+T$  ], the departure rate starts decreasing again and the latest departures occur at about  $t_{rs}+D_{rs}$ .

To analyse the way in which the time dependent set of reasonable routes is formulated, the O-D pair 4-20 was taken as an example (fig. 8.5). Thus, initially, when the prevailing traffic conditions are very close to free flow, the only routes which are considered by the travellers as reasonable choices, are 1, 2 and 3. These routes remain as feasible alternatives during the whole peak period. Routes 4, 5, 6, 7 and 8 are considered as reasonable only within specific time intervals during the peak. This is because the latter routes include links which, for certain departure times, *take the driver nearer to the origin* (section 6.2.1). Thus, for example consider the routes 4 and 5 represented by the sets  $L_{4,4-20} = \{9,21,55,52\}$  and  $L_{5,4-20} = \{9,21,56,59\}$ . These routes

include link 21 which connects node 8 to node 18. At low levels of traffic congestion, the shortest travel time from node 4 to 8 is higher than the shortest travel time to node 18. Therefore, these routes are not considered (i.e. from the definition of reasonable paths) as feasible alternatives by the drivers who depart from node 1 earlier than 8:15 or later than 9:15). After 8:15 congestion increases rapidly in routes 1, 2 and 3, mainly caused by the high demand to capacity ratio in links 42 and 44. Thus the shortest travel time to node 8 is not now longer than the one to node 18 and therefore routes including link 21 can now be considered as reasonable choices.

To demonstrate the model outputs that are related to link characteristics the evolution of link density and speed during the morning peak are depicted in figure 8.10. The links presented are 8, 40, 44, 48, 52, 55, 56 and 62. Speed is expressed in km/h and represents the average speed achieved when entering the link at a given time. Link density is not represented in its absolute definition i.e. in vehs/lane mile, but instead the number of vehicles within the link is plotted against time. The figure illustrates the capability of the model to predict the time dependent traffic conditions, and demonstrates the build up and dissipation of congestion within the network links.

The base case simulation results for the O-D pairs 2-19 and 4-20 are presented in tables 8.2a, and 8.2b respectively. The tables present the paths connecting each O-D pair that are considered as reasonable options, the demand that they attract, the average travel time, average early and late schedule delay and free flow travel time. These tables also provide the time intervals within which each of the paths is considered as reasonable (0 corresponds to 7:00 and 180 to 10:00).

Furthermore, table 8.3 provides the total and average travel time, wait time (the difference between travel time and free flow travel time), early and late schedule delay characteristic to each O-D pair and the above measures for the total demand using the network.



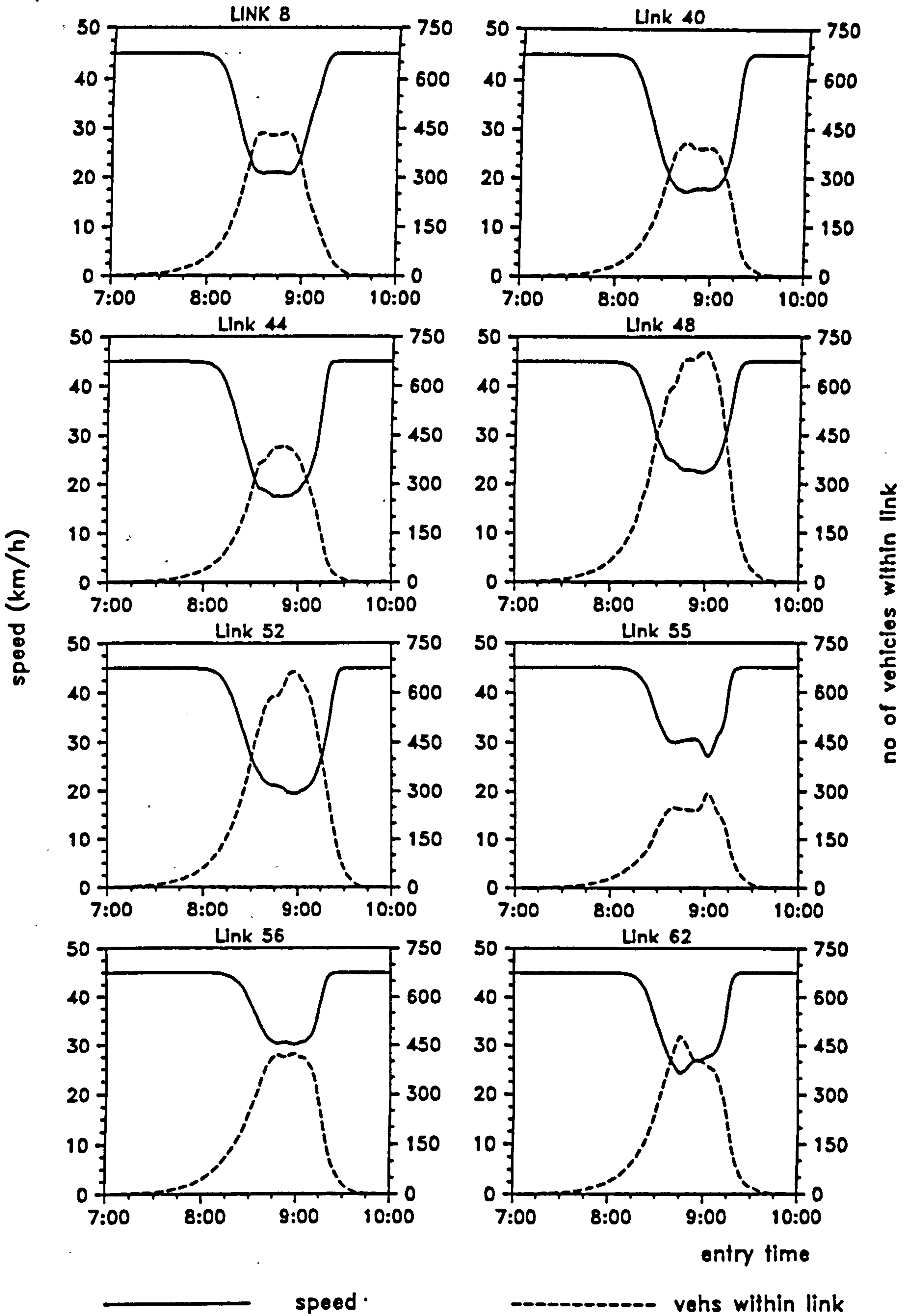


Figure 8.10: Time varying distribution of speeds and number of vehicles within various links of the network.

**Table 8.2a:** Demand and attributes of routes connecting the O-D pair 2-19, for the base case experiments.

O/D PAIR : 2 - 19		Total Demand = 2800.						
Path	Links	Demand	Av. Trav. Time	FreeFlow Time	Av. Sch. Delay early	Delay late	Average Dis.x100	
7	4 - 36 - 8 - 44 - 56 -	394.	21.4	15.3	7.5	0.7	295.3	
8	4 - 37 - 39 - 44 - 56 -	478.	19.4	14.0	7.6	0.6	272.8	
9	4 - 37 - 40 - 51 - 56 -	397.	20.5	14.7	8.5	0.6	290.4	
10	4 - 37 - 40 - 52 - 63 -	352.	20.8	14.0	9.8	0.6	301.1	
11	5 - 47 - 17 - 63 -	560.	18.8	16.1	5.8	0.5	250.2	
12	5 - 48 - 51 - 56 -	327.	22.5	17.5	8.0	0.5	305.0	
13	5 - 48 - 52 - 63 -	292.	22.9	16.8	9.1	0.5	315.6	
Total		2800.						
Average			20.6		7.8	0.6	285.7	
Time Interval	Paths							
0 - 180	7, 8, 9, 10, 11, 12, 13							

**Table 8.2b:** Demand and attributes of routes connecting the O-D pair 4-20, for the base case experiments.

O/D PAIR : 4 - 20		Total Demand = 2800.						
Path	Links	Demand	Av. Trav. Time	FreeFlow Time	Av. Sch. Delay early	Delay late	Average Dis.x100	
23	10 - 43 - 40 - 52 -	435.	18.2	12.0	8.7	0.7	268.9	
24	10 - 44 - 55 - 52 -	396.	18.9	12.7	9.0	0.7	277.8	
25	10 - 44 - 56 - 59 -	510.	17.4	13.3	7.3	0.6	247.7	
26	9 - 21 - 55 - 52 -	241.	23.0	16.1	1.0	0.8	273.0	
27	9 - 21 - 56 - 59 -	341.	20.3	16.7	1.2	0.6	240.5	
28	9 - 20 - 24 - 59 -	268.	20.9	18.9	0.0	0.6	239.0	
29	9 - 20 - 25 - 71 - 27	285.	20.6	19.1	0.0	0.6	234.1	
30	9 - 20 - 23 - 27 -	324.	19.7	18.3	0.0	0.5	223.7	
Total		2800.						
Average			19.5		4.2	0.6	251.5	
Time Interval	Paths							
0 - 75	23, 24, 25							
75 - 85	26, 27, 23, 24, 25							
85 - 130	29, 30, 28, 26, 27, 23, 24, 25							
130 - 135	26, 27, 23, 24, 25							
135 - 180	23, 24, 25							

**Table 8.3:** System's performance and trip attributes for each O-D pair in the base case experiments

PAIR	DEMAND		TRAVEL TIME	WAIT TIME	EARLY S.D.	LATE S.D.	DISUTILITY
1 - 17	2900.	TOT.(h)	933.3	203.1	339.3	29.5	7745.5
		AV.(mins)	19.3	4.2	7.0	0.6	2.671
2 - 19	2800.	TOT.(h)	962.7	243.8	364.9	27.3	7999.2
		AV.(mins)	20.6	5.2	7.8	0.6	2.857
3 - 20	2600.	TOT.(h)	849.1	283.3	319.2	33.9	7194.7
		AV.(mins)	19.6	6.5	7.4	0.8	2.767
4 - 20	2800.	TOT.(h)	910.2	192.5	195.6	29.9	7043.1
		AV.(mins)	19.5	4.1	4.2	0.6	2.515
5 - 19	2600.	TOT.(h)	891.1	112.0	205.8	24.8	6882.5
		AV.(mins)	20.6	2.6	4.7	0.6	2.647
6 - 17	2300.	TOT.(h)	727.4	134.2	207.9	19.0	5754.7
		AV.(mins)	19.0	3.5	5.4	0.5	2.502
7 - 15	2200.	TOT.(h)	654.0	146.1	218.0	20.9	5354.2
		AV.(mins)	17.8	4.0	5.9	0.6	2.434
8 - 13	2800.	TOT.(h)	679.5	134.2	217.1	23.6	5553.9
		AV.(mins)	14.6	2.9	4.7	0.5	1.984
9 - 14	2700.	TOT.(h)	787.3	149.7	262.7	25.7	6453.5
		AV.(mins)	17.5	3.3	5.8	0.6	2.390
10 - 15	2800.	TOT.(h)	845.4	165.7	226.9	28.6	6729.7
		AV.(mins)	18.1	3.6	4.9	0.6	2.403
11 - 20	2600.	TOT.(h)	995.3	269.3	281.3	36.3	8018.9
		AV.(mins)	23.0	6.2	6.5	0.8	3.084
12 - 18	2700.	TOT.(h)	905.1	212.4	288.1	26.0	7312.6
		AV.(mins)	20.1	4.7	6.4	0.6	2.708
TOTAL	31800.	TOT.(h)	10140.5	2246.3	3126.8	325.5	82042.5
		AV.(mins)	19.1	4.2	5.9	0.6	2.579

### 8.3 Sensitivity to different work start time flexibilities

In this section, the dynamic simulation model is employed to investigate the type of impacts that changes in the flexibility of the work start time can have on the pattern of congestion during the peak period. It is shown that the major effect of increasing the work start time flexibility is to reduce the level of the traffic congestion. Similar results to the ones presented here, were also obtained by Ben-Akiva et al. (1986) for a situation present in a single O-D pair network connected with parallel routes.

Two simulation experiments were conducted using work start time flexibilities 0 and 30 mins which were assumed to be the same for all the O-D pairs; the other model parameters were assumed to take their base case values. For  $D_{rs}$  equal to 0 mins, the travellers desired arrival time is at 9:00; work start time flexibility equal to 30 mins implies that travellers desire to arrive at their destinations within the period [8:30 am to 9:30 am].

To demonstrate the results, the stationary distributions of departure rates and travel times for the O-D pair 4-20 are illustrated in fig. 8.11 and 8.12 respectively. These figures depict the equilibrium distribution of departure rates and travel times for three different levels of work starting time flexibility, 0, 15 and 30 mins. In all the experiments, routes 1,2 and 3 are considered as reasonable choices during the entire commuting period, while routes 4, 5, 6 are considered as reasonable only within specific time intervals during the peak. As was expected increasing  $D_{rs}$  causes a spreading of the departure rate distribution over a longer time period, resulting in lower O-D travel times.

For  $D_{rs}$  equal to zero, the departure rate distributions exhibit a single peak in contrast to the distributions describing the other experiments where desired arrival times lie within a specific time interval of certain length. The highest departure rate is generated at the departure time that results in an on-time arrival. This is the time denoted by the variable  $t_{k,rs}^1$  which in the  $D_{rs}=0$  case coincides with the variable  $t_{k,rs}^2$ . The peak occurs since the left and right derivatives of the utility function at this departure time have opposite signs. Juxtaposition of the departure rate and travel time distributions reveals that the highest departure rate occurs just before the time that the O-D travel time attains its maximum. Departure rate distributions associated with routes 7 and 8, exhibit the highest peaks. This because, although routes 7 and 8 are associated with relatively high free flow travel times, they provide adequate capacity and therefore attract a high proportion of the demand during the most congested period.

Figure 8.12 indicates that for  $D_{rs}$  equal to 30 mins, the traffic conditions along the most of the routes connecting the O-D pair 4-20, are very close to free flow. For these

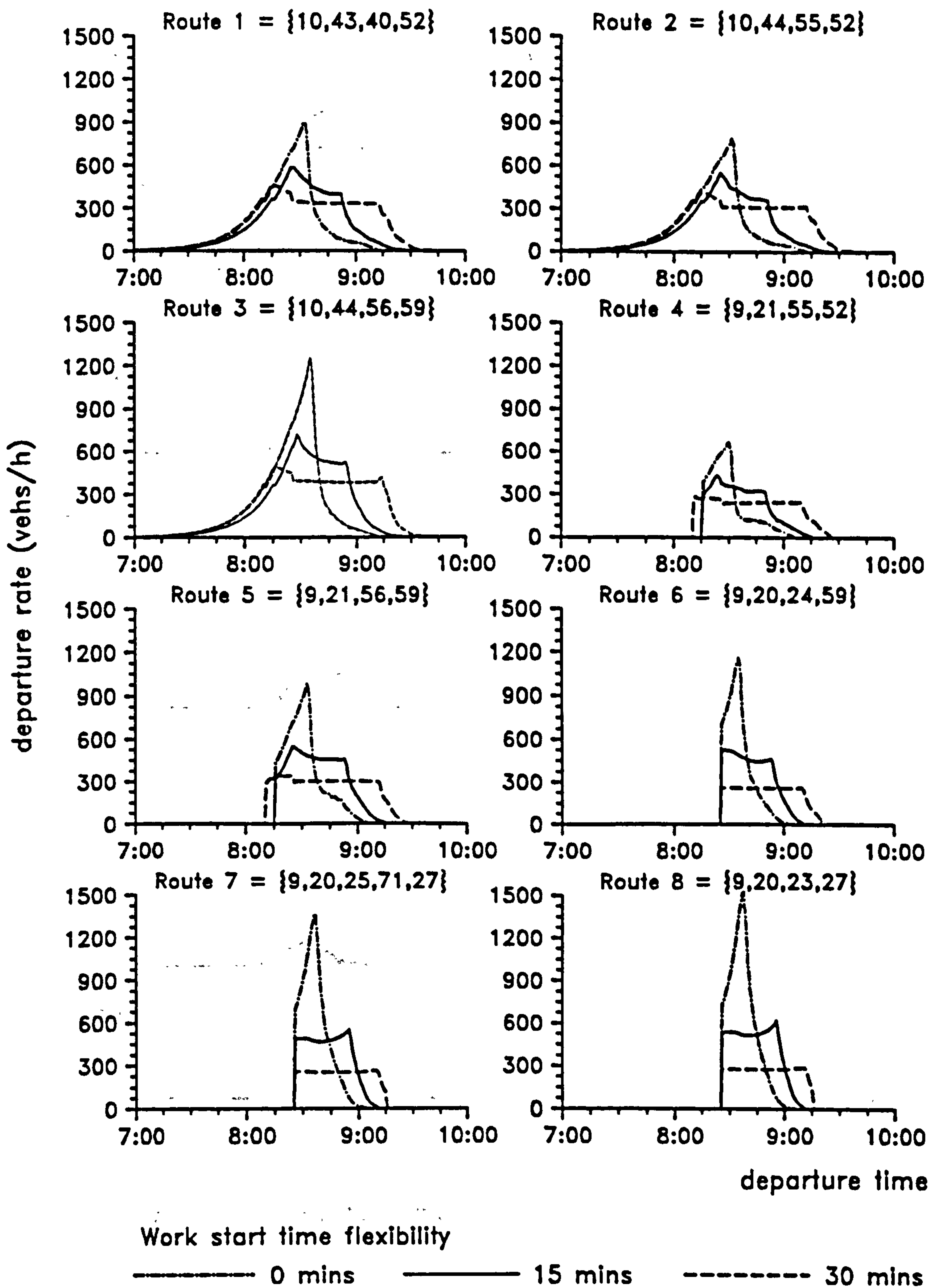


Figure 8.11: Departure rate distributions for O-D pair 4-20 under different levels of work start time flexibility.

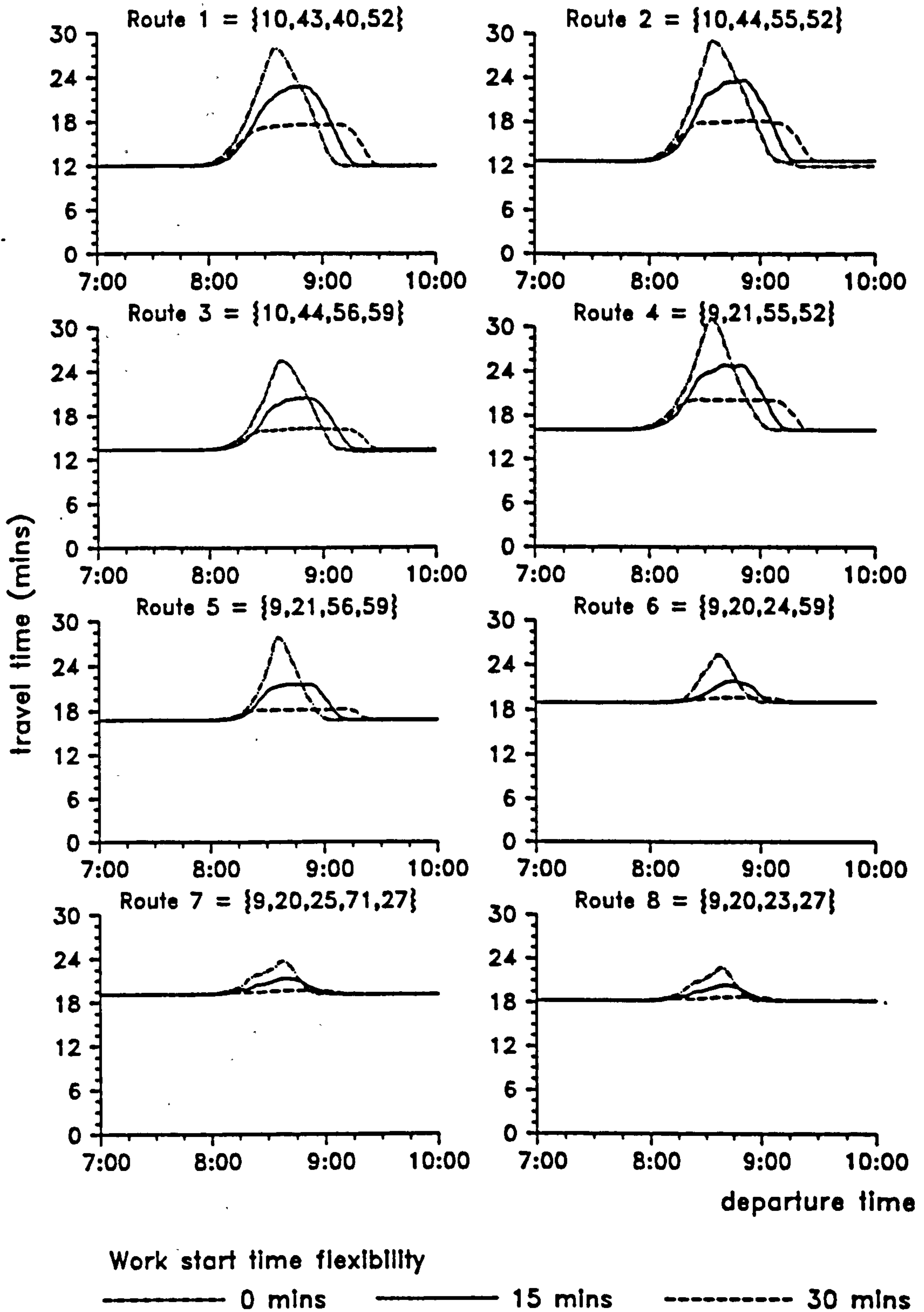


Figure 8.12: Travel time distributions for O-D pair 4-20 under different levels of work start time flexibility.

routes, departure rate distributions are almost uniform and are such that result in on-time arrivals. Departure rate distributions associated with more congested routes have primarily a uniform profile and exhibit only very low peaks. The figures also show that as the level of work start time flexibility decreases, the peaks in departure rate and travel time distributions, move towards earlier departures.

The effect that changes on work start time flexibility have on the level of congestion is also demonstrated in figure 8.13 which, for some links of the network, illustrates the average link travel times for the three different levels of  $D_{rs}$  considered in the experiments. The figure confirms the view that inflexible working times result in increasing levels of congestion due to the high concentration of traffic within a short time interval. Higher levels of work start time flexibilities result in lower levels of congestion, and therefore when possible should be introduced since they can alleviate the morning peak period congestion at a minimum cost.

Tables 8.4a and 8.5a present the demand and average values of the attributes associated with the routes connecting the O-D pair 4 - 20. Comparing these two tables and table 8.2b it appears that for  $D_{rs}=0$  there is a shift of the demand towards longer (free flow) travel time routes in order to avoid the congestion that builds up more rapidly (due to the concentration in time) in shortest routes. In the case that  $D_{rs}=30$  mins the demand is spread more evenly over time. Thus, the shortest (free flow) travel time routes do not get severely congested and offer higher level of service (even when congested) than the level offer by the uncongested longer routes. Over the whole peak, there is therefore a shift towards the shortest routes. More aggregate results for all the O-D pairs and the overall performance of the network under the two work start time flexibility scenarios are given in tables 8.4b and 8.5b.

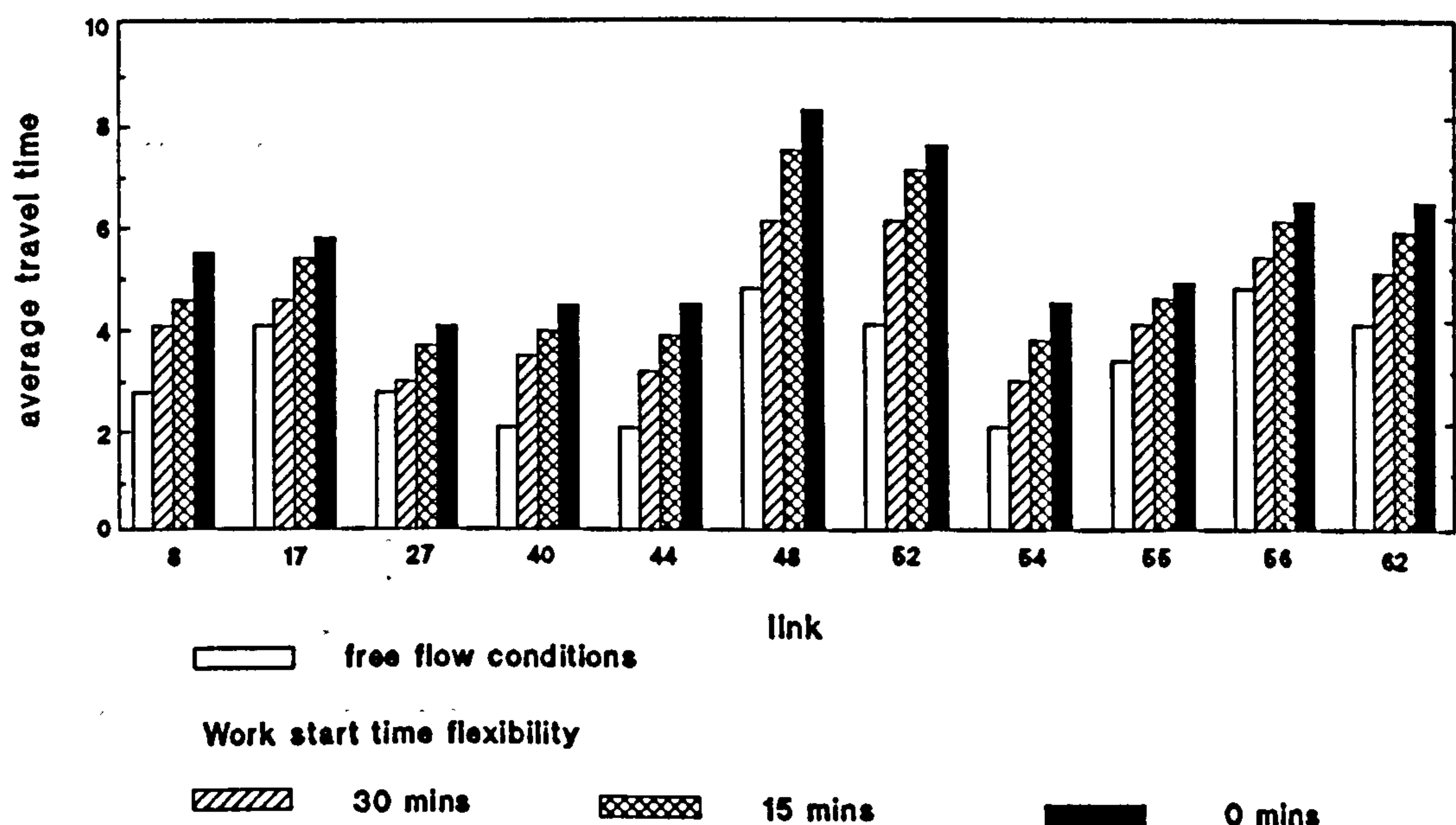


Figure 8.13 : Average link travel times for different work start time flexibilities

**Table 8.4a:** Demand and attributes of routes connecting the O-D pair 4-20, under inflexible work starting times ( $D_{rs} = 0$ ).

O/D PAIR : 4 - 20		Total Demand = 2800.					
Path	Links	Demand	Av. Trav. Time	FreeFlow Time	Av. Sch. early	Delay late	Average Dis. x100
1	10 - 43 - 40 - 52 -	420.	19.7	12.0	19.8	1.4	374.6
2	10 - 44 - 55 - 52 -	376.	20.4	12.7	20.5	1.4	385.5
3	10 - 44 - 56 - 59 -	509.	19.0	13.3	17.8	1.2	347.4
4	9 - 21 - 55 - 52 -	203.	25.5	16.1	6.9	2.3	376.3
5	9 - 21 - 56 - 59 -	311.	22.4	16.7	8.0	1.9	339.6
6	9 - 20 - 24 - 59 -	265.	23.6	18.9	2.7	2.0	321.4
7	9 - 20 - 25 - 71 - 27 -	333.	22.6	19.1	2.6	1.8	303.6
8	9 - 20 - 23 - 27 -	384.	21.6	18.3	2.8	1.7	293.2
Total		2800.					
Average			21.4	15.5	11.3	1.6	342.7
Time Interval	Paths						
0 - 75	1, 2, 3						
75 - 85	4, 5, 1, 2, 3						
85 - 125	7, 8, 6, 4, 5, 1, 2, 3						
125 - 180	1, 2, 3						

**Table 8.5a:** Demand and attributes of routes connecting the O-D pair 4-20, for 30 mins work starting time flexibility.

O/D PAIR : 4 - 20		Total Demand = 2800.					
Path	Links	Demand	Av. Trav. Time	FreeFlow Time	Av. Sch. early	Delay late	Average Dis. x100
1	10 - 43 - 40 - 52 -	509.	15.9	12.0	4.5	0.4	208.2
2	10 - 44 - 55 - 52 -	475.	16.5	12.7	4.5	0.4	214.8
3	10 - 44 - 56 - 59 -	557.	15.3	13.3	3.7	0.3	195.4
4	9 - 21 - 55 - 52 -	272.	19.8	16.1	0.0	0.4	220.5
5	9 - 21 - 56 - 59 -	339.	18.0	16.7	0.0	0.3	199.7
6	9 - 20 - 24 - 59 -	214.	19.5	18.9	0.0	0.3	215.6
7	9 - 20 - 25 - 71 - 27 -	205.	19.5	19.1	0.0	0.1	211.5
8	9 - 20 - 23 - 27 -	228.	18.7	18.3	0.0	0.1	201.4
Total		2800.					
Average			17.3	14.9	2.3	0.3	207.2
Time Interval	Paths						
0 - 70	1, 2, 3						
70 - 85	4, 5, 1, 2, 3						
85 - 135	7, 8, 6, 4, 5, 1, 2, 3						
135 - 140	6, 4, 5, 1, 2, 3						
140 - 145	4, 5, 1, 2, 3						
145 - 180	1, 2, 3						



**Table 8.4b: System's performance and trip attributes for each O-D pair for inflexible work start times.**

PAIR	DEMAND		TRAVEL TIME	WAIT TIME	EARLY S.D.	LATE S.D.	DISUTILITY
1 - 17	2900.	TOT.(h)	1051.2	318.8	698.9	75.3	10598.9
		AV.(mins)	21.7	6.6	14.5	1.6	3.655
2 - 19	2800.	TOT.(h)	1045.0	305.2	827.3	77.1	11086.7
		AV.(mins)	22.4	6.5	17.7	1.7	3.960
3 - 20	2600.	TOT.(h)	939.9	364.1	735.9	77.0	10056.0
		AV.(mins)	21.7	8.4	17.0	1.8	3.868
4 - 20	2800.	TOT.(h)	996.8	272.7	526.8	76.5	9596.8
		AV.(mins)	21.4	5.8	11.3	1.6	3.427
5 - 19	2600.	TOT.(h)	993.6	196.5	497.7	62.6	9251.6
		AV.(mins)	22.9	4.5	11.5	1.4	3.558
6 - 17	2300.	TOT.(h)	783.9	199.4	505.1	72.1	8083.1
		AV.(mins)	20.4	5.2	13.2	1.9	3.514
7 - 15	2200.	TOT.(h)	702.8	189.1	527.7	45.0	7239.9
		AV.(mins)	19.2	5.2	14.4	1.2	3.291
8 - 13	2800.	TOT.(h)	767.9	215.0	576.2	61.6	8098.5
		AV.(mins)	16.5	4.6	12.3	1.3	2.892
9 - 14	2700.	TOT.(h)	882.8	231.3	643.8	72.4	9262.5
		AV.(mins)	19.6	5.1	14.3	1.6	3.431
10 - 15	2800.	TOT.(h)	924.4	237.9	598.7	78.6	9446.9
		AV.(mins)	19.8	5.1	12.8	1.7	3.374
11 - 20	2600.	TOT.(h)	1082.6	346.2	681.8	103.7	11165.4
		AV.(mins)	25.0	8.0	15.7	2.4	4.294
12 - 18	2700.	TOT.(h)	1006.0	310.5	656.8	68.3	10038.2
		AV.(mins)	22.4	6.9	14.6	1.5	3.718
TOTAL		TOT.(h)	11176.9	3186.7	7476.7	870.2	113924.5
		AV.(mins)	21.1	6.0	14.1	1.6	3.5825

**Table 8.5b:** System's performance and trip attributes for each O-D pair for 30 mins work start time flexibility.

PAIR	DEMAND		TRAVEL TIME	WAIT TIME	EARLY S.D.	LATE S.D.	DISUTILITY
1 - 17	2900.	TOT.(h)	844.8	109.8	165.6	12.9	6249.0
		AV.(mins)	17.5	2.3	3.4	0.3	2.155
2 - 19	2800.	TOT.(h)	849.1	131.6	166.7	14.1	6298.8
		AV.(mins)	18.2	2.8	3.6	0.3	2.250
3 - 20	2600.	TOT.(h)	759.2	173.5	143.6	17.1	5679.1
		AV.(mins)	17.5	4.0	3.3	0.4	2.184
4 - 20	2800.	TOT.(h)	805.9	110.1	108.9	14.4	5801.5
		AV.(mins)	17.3	2.4	2.3	0.3	2.072
5 - 19	2600.	TOT.(h)	842.8	62.5	93.9	10.7	5922.8
		AV.(mins)	19.4	1.4	2.2	0.2	2.278
6 - 17	2300.	TOT.(h)	614.8	54.8	119.4	9.0	4538.1
		AV.(mins)	16.0	1.4	3.1	0.2	1.973
7 - 15	2200.	TOT.(h)	572.8	75.4	114.0	9.6	4256.3
		AV.(mins)	15.6	2.1	3.1	0.3	1.935
8 - 13	2800.	TOT.(h)	596.9	58.1	118.8	11.3	4454.6
		AV.(mins)	12.8	1.2	2.5	0.2	1.591
9 - 14	2700.	TOT.(h)	699.9	67.5	134.6	11.0	5172.1
		AV.(mins)	15.6	1.5	3.0	0.2	1.916
10 - 15	2800.	TOT.(h)	717.9	77.1	128.4	12.1	5279.8
		AV.(mins)	15.4	1.7	2.8	0.3	1.886
11 - 20	2600.	TOT.(h)	858.3	182.2	146.5	17.1	6324.0
		AV.(mins)	19.8	4.2	3.4	0.4	2.432
12 - 18	2700.	TOT.(h)	804.6	110.6	147.7	11.7	5903.2
		AV.(mins)	17.9	2.5	3.3	0.3	2.186
TOTAL	31800.	TOT.(h)	8967.0	1213.2	1588.1	151.0	65879.3
		AV.(mins)	16.9	2.3	3.0	0.3	2.07167

#### 8.4 Sensitivity to different levels of variability of preferences

In this section the sensitivity of the model outputs to changes in the parameters reflecting the travellers' variability of preferences with respect to trip choices, is investigated.

The sensitivity to changes in the parameters  $\mu_r$  and  $\mu_t$ , which reflect travellers' variability of preferences with respect to route and departure time choice, is investigated through the results of two simulation experiments. The outputs of these experiments are compared to the results derived from the base case experiment. In the first experiment, both  $\mu_r$  and  $\mu_t$  have taken the value 0.5 reflecting a high variability in travellers' preferences when compared to the one considered as the base case scenario. Increased variability is attributable to perception errors due to lack of information on the characteristics of the alternative combinations of route and departure time, probabilistic decision making and/or particular preferences of the individuals. Within this "more" stochastic choice context, there is a greater percentage of travellers who believe that will maximise their utility of travel by selecting alternatives associated with lower (observed) utility levels. Figure 8.14 illustrates the departure rate distributions for the three different levels of variability of preferences. As was expected, for  $\mu_r = \mu_t = 0.5$ , the departure rate distributions exhibit lower peaks compared to the base case distributions and are spread over a longer period of time. This, as illustrated in figure 8.15 results in lower levels of congestion and for some of the routes in almost free flow conditions.

In the second experiment both  $\mu_r$  and  $\mu_t$  were assumed to take the value 2.0 . Under this assumption, travellers have a better perception of the utility associated with each combination of route and departure time choice. Furthermore, since they want to maximise their utility of travel their choices are concentrated in the vicinity of the highest utility alternatives. Thus, within this more deterministic choice context, departure rate distributions are less spread and exhibit high peaks. As shown in figure 8.15, this high concentration of departures around the optimum choices results in higher levels of congestion and longer travel times, (see tables 8.6 and 8.7).

Numeric outputs of the simulation runs are presented in tables 8.6 and 8.7. The tables indicate that high variability of preferences results in lower levels of congestion and lower travel times but higher levels of schedule delay. In the first experiments the total travel time for all the users in the network is 9084 hours, i.e. 1054 hours lower than the total travel time spent in the base case scenario. This reflects the tendency of a higher proportion of travellers to select "non optimal" departure times, and thus

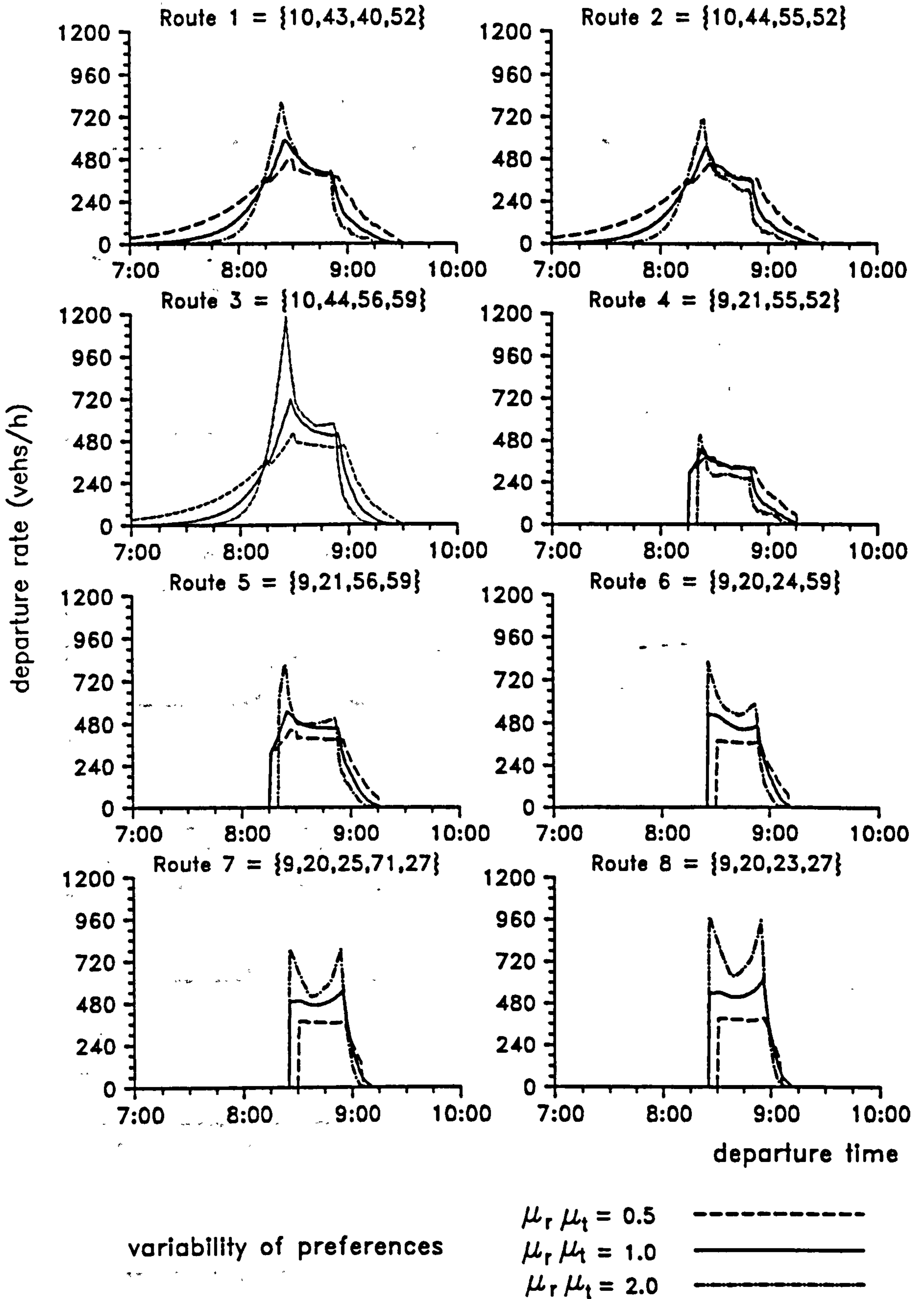


Figure 8.14: Departure rate distributions for O-D pair 4-20 under different levels of drivers' variability of preferences.

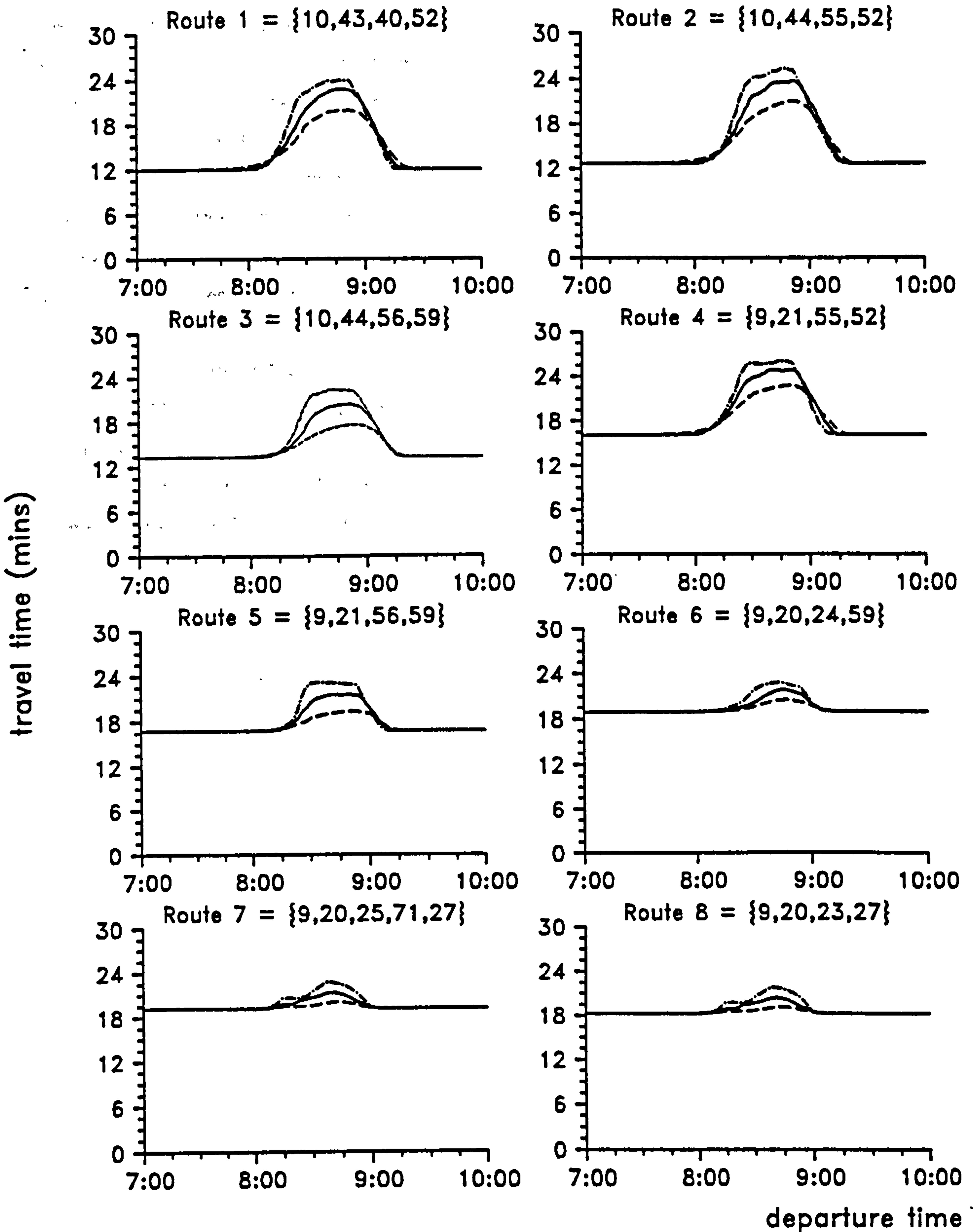


Figure 8.15: Travel time distributions for O-D pair 4-20 under different levels of drivers' variability of preferences.

demand is distributed more uniformly and over a longer period implying that travellers experience lower levels of congestion but longer schedule delays; total early schedule in the first experiment was found to be 85% higher than in the base case scenario and total late schedule delay higher by 70%. On the other hand, the level of congestion occurring within the "deterministic choice" setting is higher than in the base case. This is a result of the high concentration of departures within "optimal" time intervals during the peak and implies that travellers with low variability of preferences tend to arrive at their destination on time at the expense of a higher travel times. The total travel time is 11014 hours, i.e. 8.6% higher than its corresponding base case value, and total early and late delays are 1745 and 163 hours respectively compared to the 3126 hours and 325 hours found in the base case experiment. Thus, although travellers with lower variability of preferences experience longer travel times, on the whole they are better off since their trip choices result in low levels of schedule delay which result in higher levels of utility of travel.

Table 8.6: System's performance and trip attributes for each O-D pair, under low levels of travellers' variability of preferences;  $\mu_r, \mu_t = 2.0$

PAIR	DEMAND		TRAVEL TIME	WAIT TIME	EARLY S.D.	LATE S.D.	DISUTILITY
1 - 17	2900.	TOT.(h)	1026.2	298.4	142.8	17.9	7396.2
		AV.(mins)	21.2	6.2	3.0	0.4	2.550
2 - 19	2800.	TOT.(h)	1064.0	301.6	219.9	13.6	7873.2
		AV.(mins)	22.8	6.5	4.7	0.3	2.812
3 - 20	2600.	TOT.(h)	920.9	350.4	221.6	15.7	6996.6
		AV.(mins)	21.3	8.1	5.1	0.4	2.691
4 - 20	2800.	TOT.(h)	981.5	250.9	109.1	13.2	6907.9
		AV.(mins)	21.0	5.4	2.3	0.3	2.467
5 - 19	2600.	TOT.(h)	959.2	180.8	97.4	10.9	6684.8
		AV.(mins)	22.1	4.2	2.2	0.3	2.571
6 - 17	2300.	TOT.(h)	779.1	189.5	99.6	11.3	5546.8
		AV.(mins)	20.3	4.9	2.6	0.3	2.412
7 - 15	2200.	TOT.(h)	713.5	193.2	121.7	9.8	5189.6
		AV.(mins)	19.5	5.3	3.3	0.3	2.359
8 - 13	2800.	TOT.(h)	743.3	193.9	113.3	10.6	5360.4
		AV.(mins)	15.9	4.2	2.4	0.2	1.914
9 - 14	2700.	TOT.(h)	873.1	227.7	142.1	12.8	6336.8
		AV.(mins)	19.4	5.1	3.2	0.3	2.347
10 - 15	2800.	TOT.(h)	913.5	227.1	128.6	11.8	6527.5
		AV.(mins)	19.6	4.9	2.8	0.3	2.331
11 - 20	2600.	TOT.(h)	1058.2	342.0	214.2	18.5	7889.5
		AV.(mins)	24.4	7.9	4.9	0.4	3.034
12 - 18	2700.	TOT.(h)	982.3	293.0	135.1	16.8	7068.2
		AV.(mins)	21.8	6.5	3.0	0.4	2.618
TOTAL	31800.	TOT.(h)	11014.8	3048.5	1745.4	162.9	79777.5
		AV.(mins)	20.8	5.8	3.3	0.3	2.5057

**Table 8.7:** System's performance and trip attributes for each O-D pair, under high levels of travellers' variability of preferences;  $\mu_r, \mu_t = 0.5$

PAIR	DEMAND		TRAVEL TIME	WAIT TIME	EARLY S.D.	LATE S.D.	DISUTILITY
1 - 17	2900.	TOT.(h)	849.9	113.0	601.2	46.9	8497.8
		AV.(mins)	17.6	2.3	12.4	1.0	2.930
2 - 19	2800.	TOT.(h)	861.4	141.9	588.9	51.6	8595.2
		AV.(mins)	18.5	3.0	12.6	1.1	3.070
3 - 20	2600.	TOT.(h)	741.2	178.0	526.1	52.3	7590.8
		AV.(mins)	17.1	4.1	12.1	1.2	2.920
4 - 20	2800.	TOT.(h)	820.4	130.6	402.3	53.0	7625.1
		AV.(mins)	17.6	2.8	8.6	1.1	2.723
5 - 19	2600.	TOT.(h)	865.1	84.9	352.5	44.2	7583.5
		AV.(mins)	20.0	2.0	8.1	1.0	2.917
6 - 17	2300.	TOT.(h)	625.1	64.2	447.7	35.2	6282.2
		AV.(mins)	16.3	1.7	11.7	0.9	2.731
7 - 15	2200.	TOT.(h)	584.3	83.5	408.6	36.7	5890.8
		AV.(mins)	15.9	2.3	11.1	1.0	2.678
8 - 13	2800.	TOT.(h)	610.3	69.5	453.9	41.3	6304.5
		AV.(mins)	13.1	1.5	9.7	0.9	2.252
9 - 14	2700.	TOT.(h)	708.7	76.0	507.8	42.1	7156.4
		AV.(mins)	15.7	1.7	11.3	0.9	2.651
10 - 15	2800.	TOT.(h)	737.5	89.5	485.3	46.4	7317.8
		AV.(mins)	15.8	1.9	10.4	1.0	2.614
11 - 20	2600.	TOT.(h)	865.2	167.4	489.2	61.3	8376.8
		AV.(mins)	20.0	3.9	11.3	1.4	3.222
12 - 18	2700.	TOT.(h)	814.9	118.8	532.1	44.0	7959.5
		AV.(mins)	18.1	2.6	11.8	1.0	2.948
TOTAL	31800.	TOT.(h)	9084.0	1317.3	5795.6	555.0	89180.4
		AV.(mins)	17.1	2.5	10.9	1.0	2.8044



### 8.5 Sensitivity to different levels of demand

In this section two simulation experiments are conducted to explore the impacts of changes in the level of demand for travel. In the first the level of demand for each O-D pair was increased by 25% and in the second by 50%.

The equilibrium distributions of departure rates and travel times, for these two experiments are presented in figures 8.16 and 8.17. As expected, the increase in overall demand causes severe congestion problems on all the already congested routes. The shape of the distributions remain the same as in the base case scenario, but the new distributions are more highly peaked. The figures also show that as the level of demand increases the mode of the departure rate distribution shifts to the left, revealing evidence that travellers respond to the higher level of congestion, resulting from the increased level of demand and the limited network capacity, by shifting towards earlier departure times. The travel time distributions illustrate the increase in traffic congestion resulting from the increased level of demand. Peak period travel times are substantially higher and the congestion period is extended over a longer period towards earlier departures.

Both experiments show that the less congested (in the base case scenario) routes attract the highest proportion of the increased demand and vice versa. Table 8.2b (base case pair 4) indicates that routes 1 and 2 are the most congested ones since the difference between their average and free flow travel times are the highest and for both of them equal to 6.2 mins. On the other hand, routes 7 and 8 are associated with relatively high free flow travel, but they impose short delays (on average 1.5 and 1.4 mins respectively) and they are the less congested.

Comparison of tables 8.8a and 8.9a with 8.2b reveals that in the 25% demand increase scenario, routes 1 and 2 attract only 17% and 9% respectively more than the base case total number of travellers while the less congested routes 7 and 8 attract 41% and 56% respectively, i.e. on average about twice the percentage of the total demand increase. It seems that the latter routes attract a large number of travellers because of their unused excess capacity. Similarly in the 50% demand increase experiment, routes 1 and 2 attract 37% and 30% respectively, while for routes 7 and 8 the corresponding increase was 62% and 64% respectively.

The calculated network's measures of performance reveal a large increase in total travel time and schedule delay occurs under the increased level of demand. Total travel time has increased from the base case figure of 10104 hours to 14362 hours and to 19412

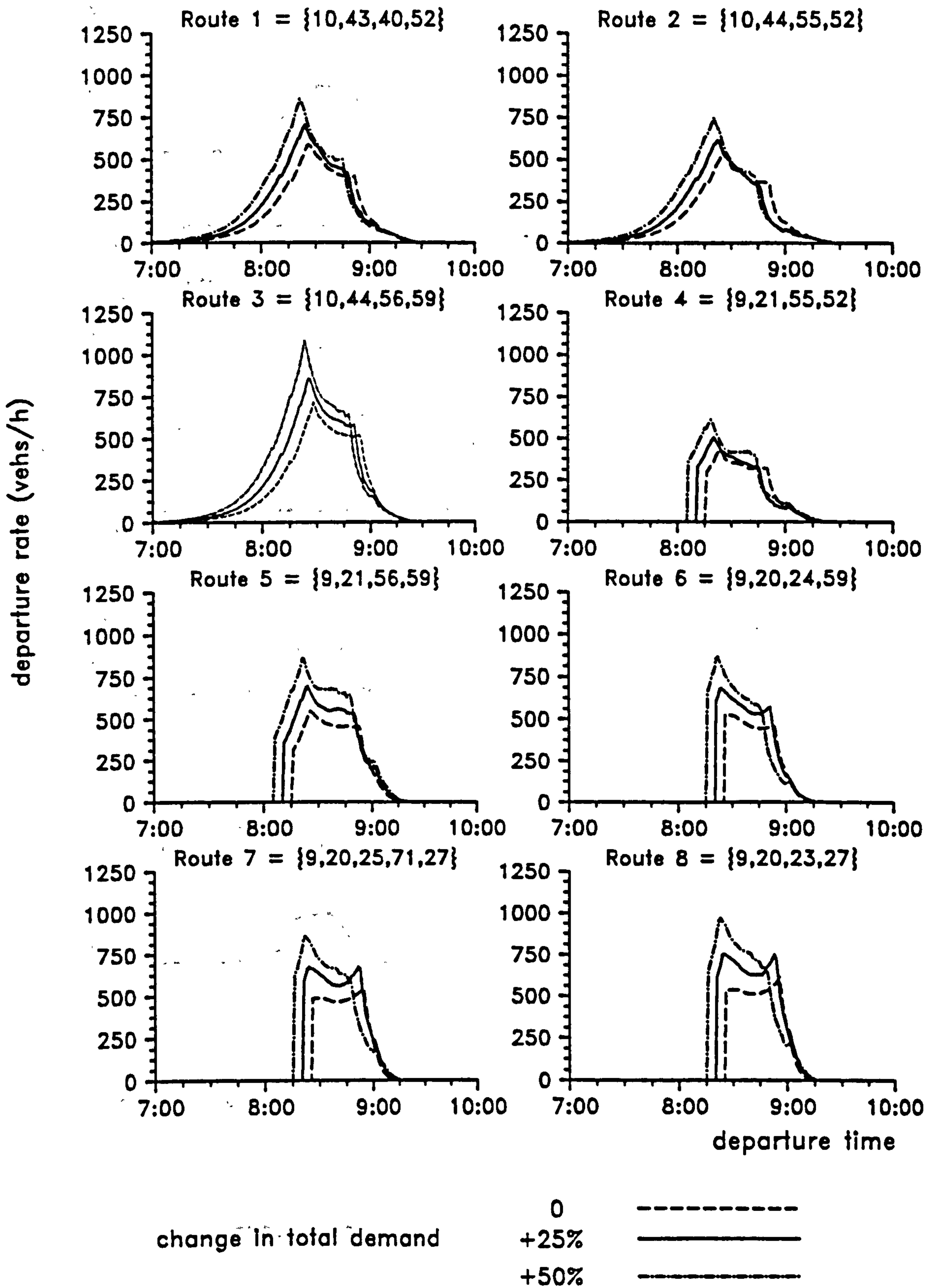


Figure 8.16: Departure rate distributions for O-D pair 4-20 under different levels of total demand.

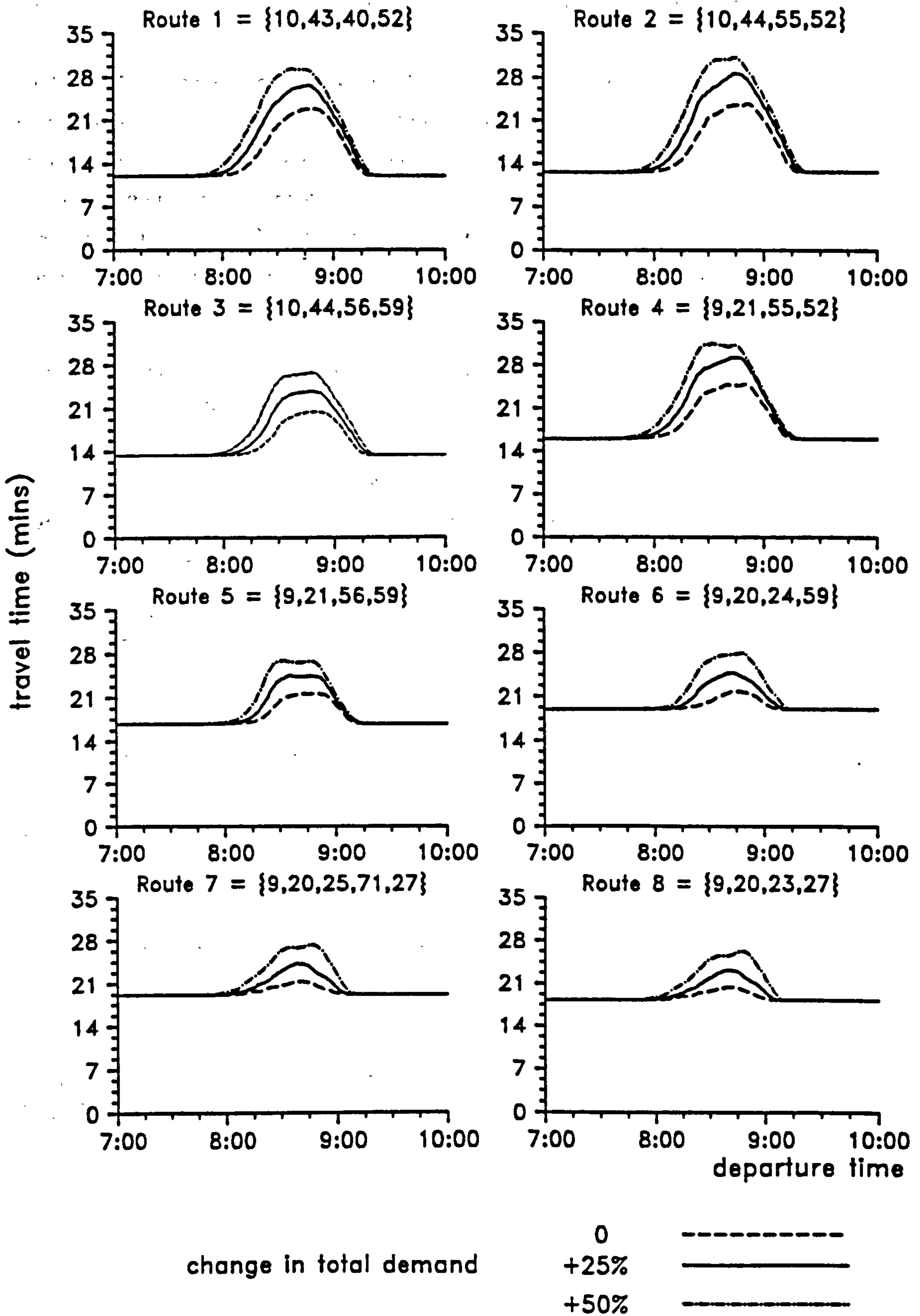


Figure 8.17: Travel time distributions for O-D pair 4-20 under different levels of total demand.

hours for the +25% and the +50% scenarios. The corresponding values of total early schedule delay are 3126h, 4389h and 5988h, and total late schedule delay 325h, 528h and 738h. This large increase is mainly due to the higher number of travellers using the network, rather than to the resulting increased level of congestion. This is evident from the increase in the average travel time and schedule delay per traveller. In the first experiment the average travel time was increased by 13.6% and the average schedule delay by 11.8%, and in the second by 27.8% and 27.1% respectively. The above figures are extracted from tables 8.8b and 8.9b.

**Table 8.8a:** Demand and attributes of routes connecting the O-D pair 4-20, under 25% increase in the level of demand.

O/D PAIR : 4 - 20		Total Demand = 3500.					
Path	Links	Demand	Av. Trav. Time	FreeFlow Time	Av. Sch. Delay early	Delay late	Average Dis.x100
39	10 - 43 - 40 - 52 -	509.	20.4	12.0	10.1	0.8	302.3
40	10 - 44 - 55 - 52 -	434.	21.2	12.7	11.0	0.8	316.6
41	10 - 44 - 56 - 59 -	615.	19.4	13.3	8.6	0.6	278.9
42	9 - 21 - 55 - 52 -	276.	25.7	16.1	1.8	1.0	312.9
43	9 - 21 - 56 - 59 -	447.	22.3	16.7	2.1	0.8	271.4
44	9 - 20 - 24 - 59 -	362.	23.3	18.9	0.2	0.7	268.1
45	9 - 20 - 25 - 71 - 27 -	401.	22.7	19.1	0.1	0.7	261.4
46	9 - 20 - 23 - 27 -	456.	21.7	18.3	0.2	0.7	250.5
Total		3500.					
Average			21.7	15.6	4.8	0.8	281.9
Time Interval	Paths						
0 - 70	39, 40, 41						
70 - 80	42, 43, 39, 40, 41						
80 - 135	45, 46, 44, 42, 43, 39, 40, 41						
135 - 145	42, 43, 39, 40, 41						

**Table 8.8b: System's performance and trip attributes for each O-D pair under 25% increase in the level of demand.**

PAIR	DEMAND		TRAVEL TIME	WAIT TIME	EARLY S.D.	LATE S.D.	DISUTILITY
1 - 17	3625.	TOT.(h)	1324.1	394.1	443.3	47.9	10932.1
		AV.(mins)	21.9	6.5	7.3	0.8	3.016
2 - 19	3500.	TOT.(h)	1367.3	406.5	502.6	52.1	11502.6
		AV.(mins)	23.4	7.0	8.6	0.9	3.286
3 - 20	3250.	TOT.(h)	1183.2	480.1	477.2	50.0	10193.5
		AV.(mins)	21.8	8.9	8.8	0.9	3.136
4 - 20	3500.	TOT.(h)	1266.4	356.8	280.2	43.9	9865.6
		AV.(mins)	21.7	6.1	4.8	0.8	2.816
5 - 19	3250.	TOT.(h)	1265.5	263.7	275.8	36.3	9727.0
		AV.(mins)	23.4	4.9	5.1	0.7	2.993
6 - 17	2875.	TOT.(h)	1065.0	282.2	274.9	40.7	8506.5
		AV.(mins)	22.2	5.9	5.7	0.8	2.959
7 - 15	2750.	TOT.(h)	922.3	264.4	299.3	30.6	7535.7
		AV.(mins)	20.1	5.8	6.5	0.7	2.740
8 - 13	3500.	TOT.(h)	960.5	271.7	311.1	33.4	7869.0
		AV.(mins)	16.5	4.7	5.3	0.6	2.248
9 - 14	3375.	TOT.(h)	1148.1	316.5	371.1	41.1	9420.8
		AV.(mins)	20.4	5.6	6.6	0.7	2.791
10 - 15	3500.	TOT.(h)	1190.6	325.7	339.6	43.6	9607.7
		AV.(mins)	20.4	5.6	5.8	0.7	2.745
11 - 20	3250.	TOT.(h)	1403.5	447.1	393.6	63.6	11484.7
		AV.(mins)	25.9	8.3	7.3	1.2	3.534
12 - 18	3375.	TOT.(h)	1265.8	384.8	420.5	45.0	10425.8
		AV.(mins)	22.5	6.8	7.5	0.8	3.089
TOTAL	39750.	TOT.(h)	14362.3	4193.6	4389.2	528.2	117071.0
		AV.(mins)	21.7	6.3	6.6	0.8	2.94518

**Table 8.9a:** Demand and attributes of routes connecting the O-D pair 4-20, under 50% increase in the level of demand.

O/D PAIR : 4 - 20		Total Demand = 4200.					
Path	Links	Demand	Av. Trav. Time	FreeFlow Time	Av. Sch. Delay early	Delay late	Average Dis.x100
56	10 - 43 - 40 - 52 -	599.	22.2	12.0	11.3	0.8	330.9
57	10 - 44 - 55 - 52 -	515.	23.0	12.7	12.2	0.8	344.5
58	10 - 44 - 56 - 59 -	730.	21.2	13.3	9.8	0.7	306.8
59	9 - 21 - 55 - 52 -	358.	27.6	16.1	2.7	1.2	341.1
60	9 - 21 - 56 - 59 -	577.	24.0	16.7	2.9	0.9	299.0
61	9 - 20 - 24 - 59 -	425.	25.8	18.9	0.7	0.9	300.8
62	9 - 20 - 25 - 71 - 27 -	463.	25.4	19.1	0.5	0.9	297.9
63	9 - 20 - 23 - 27 -	534.	24.2	18.3	0.7	0.9	285.3
Total		4200.					
Average			23.8	15.6	5.6	0.9	312.4
Time Interval	Paths						
0 - 65	56, 57, 58						
65 - 75	59, 60, 56, 57, 58						
75 - 135	62, 63, 61, 59, 60, 56, 57, 58						
135 - 145	61, 59, 60, 56, 57, 58						

**Table 8.9b:** System's performance and trip attributes for each O-D pair under 50% increase in the level of demand.

PAIR	DEMAND		TRAVEL TIME	WAIT TIME	EARLY S.D.	LATE S.D.	DISUTILITY
1 - 17	4350.	TOT.(h)	1801.0	669.6	618.9	70.2	15007.1
		AV.(mins)	24.8	9.2	8.5	1.0	3.450
2 - 19	4200.	TOT.(h)	1881.0	688.8	660.8	71.7	15706.1
		AV.(mins)	26.9	9.8	9.4	1.0	3.740
3 - 20	3900.	TOT.(h)	1683.0	735.7	578.5	73.8	14149.8
		AV.(mins)	25.9	11.3	8.9	1.1	3.628
4 - 20	4200.	TOT.(h)	1667.4	575.8	395.0	59.8	13120.9
		AV.(mins)	23.8	8.2	5.6	0.9	3.124
5 - 19	3900.	TOT.(h)	1692.3	478.3	427.9	63.9	13471.4
		AV.(mins)	26.0	7.4	6.6	1.0	3.454
6 - 17	3450.	TOT.(h)	1370.7	466.0	429.1	57.6	11322.7
		AV.(mins)	23.8	8.1	7.5	1.0	3.282
7 - 15	3300.	TOT.(h)	1228.5	440.1	420.9	37.3	10071.2
		AV.(mins)	22.3	8.0	7.7	0.7	3.052
8 - 13	4200.	TOT.(h)	1285.4	456.6	423.5	48.1	10610.4
		AV.(mins)	18.4	6.5	6.1	0.7	2.526
9 - 14	4050.	TOT.(h)	1558.1	516.6	479.8	57.0	12710.7
		AV.(mins)	23.1	7.7	7.1	0.8	3.138
10 - 15	4200.	TOT.(h)	1585.5	523.1	466.6	52.1	12758.8
		AV.(mins)	22.6	7.5	6.7	0.7	3.038
11 - 20	3900.	TOT.(h)	1931.0	742.5	518.7	85.4	15679.5
		AV.(mins)	29.7	11.4	8.0	1.3	4.020
12 - 18	4050.	TOT.(h)	1728.3	637.1	568.6	61.8	14218.2
		AV.(mins)	25.6	9.4	8.4	0.9	3.511
TOTAL	47700.	TOT.(h)	19412.2	6930.2	5988.3	738.7	158826.8
		AV.(mins)	24.4	8.7	7.5	0.9	3.3297

### 8.6 Simulation experiments using the queueing theory model

The commuting context considered in the simulation experiments presented in this section, in order to analyse the outputs of the DSUE model which employs the queueing theory based travel time model, consists of the network used in Hammerslag (1988). The network consists of 10 nodes and 22 links and is depicted in figure 8.18; commuters are assumed to travel between 9 O-D pairs; the total number of trips between each O-D pair is given in table 8.10. The values of the parameters used in the utility functions are the same as the ones used in the base case simulation experiment using the traffic flow based travel time model. As stated in section 6.3.4 and following Vickrey (1969) and de Palma *et al* (1983) roads are modelled as bottlenecks through which traffic flow is either uncongested or fixed at a capacity independent of traffic density.

The stationary distributions of the arrival rates at the bottlenecks in some of the links in the network are illustrated in figure 8.19. Link 3 is the most severely congested since it has a capacity of 4600 veh/h and has to accommodate all the demand originating from nodes 1 and 2, consisting of 6350 drivers who want to arrive at their destination during the period 8:45 to 9:15. The average travel time required to traverse this link during the morning peak is 9.4 mins implying a waiting time of 6.4 mins. Similarly, link 1 gets also congested, but at a lower level than link 3, since with a capacity of 3800 vehs/h it has to accommodate the demand originating from node 1 consisting of 3950 travellers. As a result of the network topography and the link capacities used in the simulation experiments, links 12 and 13, located downstream of 3, are not congested since the traffic is held at link 3. Thus, although the capacity of link 13 is 3900vehs/h and the total demand traversing this link is 4370 vehs, there is almost no built up of queues since the maximum arrival rate at this bottleneck is only slightly higher than its capacity. Link 6 is a "key" link of the network since it is used by travellers belonging to any O-D pair. Due to its limited capacity (2700 vehs/h) compared with the links upstream (link 7 for trips originating from nodes 4 and 3, and link 12 and further upstream link 3 for trips originating from nodes 1 and 2), it becomes severely congested. Link 6 holds the traffic and therefore the arrival rate at link 18 does not exceed the capacity of the latter link. Thus although the demand traversing link 18 consists of 4130 vehs and its capacity is 3350 vehs/h there is no build up of queues.

The model outcomes demonstrate the ability of the DSUE model to represent the time dimension of traffic in contrast to static assignment procedures which in the experiment presented in this section would have calculated that a certain level of



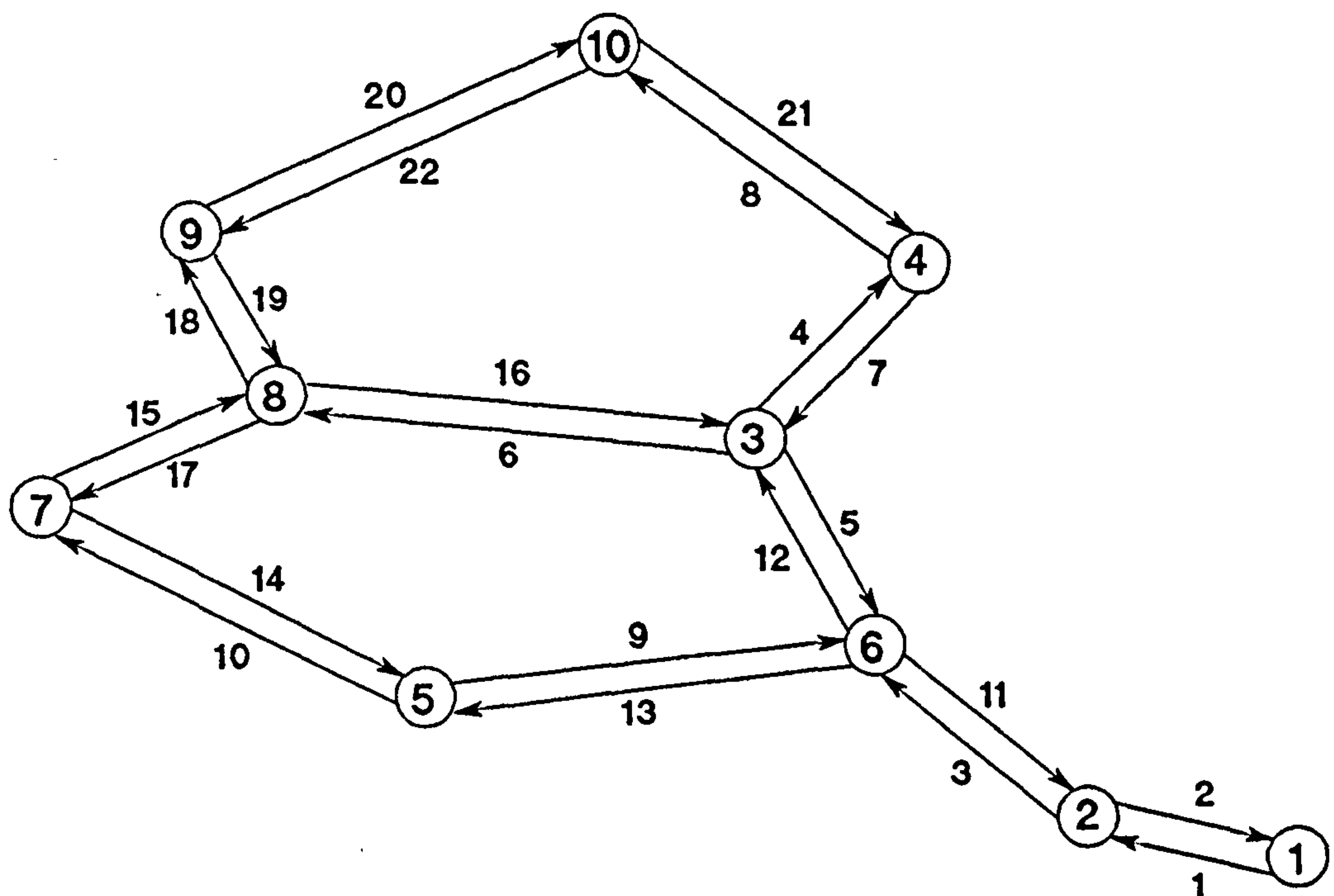


Figure 8.18 : The network used in the queuing theory based experiment

Table 8.10 : Origin - Destination pairs and demand.

O-D PAIR	DEMAND
1 - 7	1300.
1 - 9	1400.
1 - 10	1250.
2 - 9	1200.
2 - 10	1200.
3 - 7	1350.
3 - 9	1250.
4 - 7	1200.
4 - 9	1500.

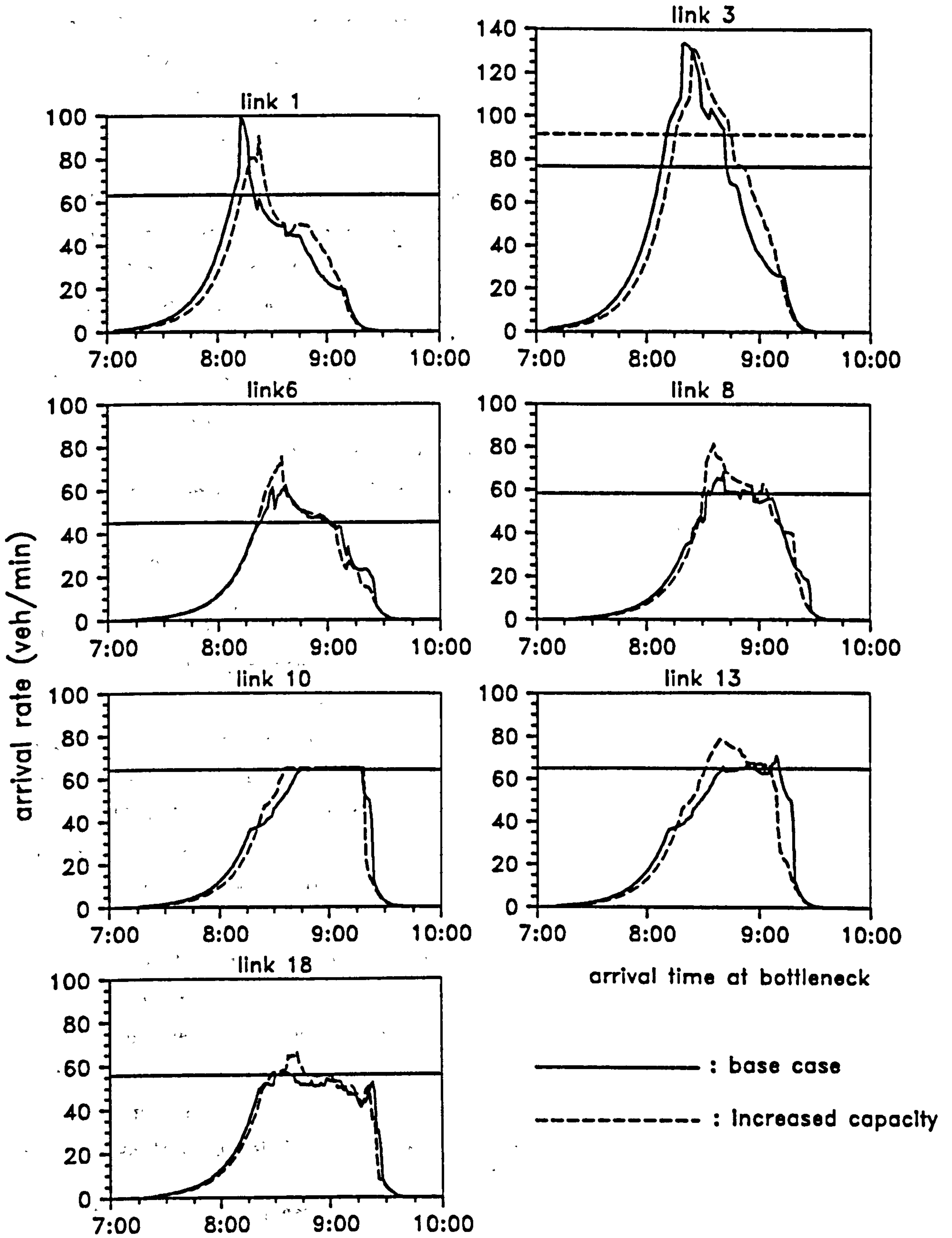


Figure 8.19: The time dependent pattern of arrivals at various bottlenecks, before and after increasing the capacity of bottleneck three.

congestion (depending on the travel time model they adopt) would develop for example in link 13. This is a result of the implicit assumption made in the static assignment models that *"cars are present on all links at the same time. So, cars which in reality are held in a bottleneck can in the calculation cause a congestion downstream"*, (Hammerslag, 1988).

The DSUE assignment model can be used to analyse the possible effects that an increase in the network capacity may have on traffic patterns and congestion levels. Thus to illustrate these effects, a second simulation experiment was conducted using the queueing theory based travel time model. In this experiment the capacity of link 3 was increased from 4600 vehs/h to 5500 vehs/hour. The capacity increase has resulted in lower levels of congestion in link 3; the average travel time was estimated to be reduced by 4 mins from 9.4 to 5.5 mins, implying that the average delay in the queue formed in bottleneck 3 is 2.5 mins. The lower levels of delay at link 3 have resulted in a shift of the demand emanating from nodes 1 and 2 towards later departures. This is illustrated by the shift of the arrival rate distributions at bottlenecks 1 and 3, as shown in figure 8.19.

Another effect that was expected is the potential increase of congestion levels in links downstream of the expanded capacity bottleneck. The DSUE model has demonstrated its ability to forecast such effects which cannot be estimated by static assignment models. The capacity expansion scheme implies that there is a lower level of traffic held at link 3. Thus, there is a higher outflow from link 3 and therefore higher inflows to links 12 and 13. Link 12 is still uncongested but the outflow from this link is now higher and since a high percentage of this outflow enters link 6, congestion in this link is increasing. The average travel time in link 6 was increased by 1.6 mins from 10 to 11.6 implying an average waiting time of 4.6 mins. Expansion of the capacity in link 3 has also created congestion in link 13, in which the average delay was increased by 1.4 mins. The higher levels of congestion developed in links 6 and 13 have resulted in a diversion of the traffic towards routes not including these links. Thus for example a higher proportion of the demand between nodes 4 and 7 follows the path {8, 22, 19, 17}. This in turn has increased the congestion level in link 8; at the equilibrium state the average delay in link 8 was increased by 2.1 mins from 0.4 to 2.5 mins.

The effects that the changes in traffic patterns and link travel times have on the O-D travel times are illustrated in figure 8.20. Average O-D travel times were not reduced for all the O-D pairs. On the contrary, for some of the O-D pairs, average travel times are higher after expanding the capacity of link 3. Average travel time from node 1 to

node 7 was reduced by 2.8 mins since travel time in link 3 was reduced by 4 mins but in link 6 it was increased by 1.6 mins. Similar average O-D travel time for the pair 1 - 9 was reduced by 1.4 mins and for the pair 1 - 10 by 2.4 mins. Drivers travelling from node 3 to 7, and from node 3 to 9 experience on average worse traffic conditions. This because the routes connecting these pairs include links 6 and 13. Average travel time for the pairs 3 - 7 was increased by 1.8 mins and for the pair 3 - 9 by 2.3 mins.

In general figure 8.20 demonstrates that drivers originating from node 1 and 2 with destinations at nodes 7, 9 and 10 are better off under the capacity expansion scheme since this has resulted in a reduction of congestion levels at link 3. However the benefits in terms of time savings is not as high as waiting time savings from link 3 since downstream links like 6 and 13 became more congested. On the other hand drivers starting their journey from nodes 3 and 4 with destinations at nodes 7, 8 and 9 experience higher levels of congestion. This since the routes connecting the O-D pairs 3-7, 3-9, 4-7 and 4-9 do not include any links in which congestion was reduced; on the contrary they include either link 6 or 13 which are characterised by higher levels of congestion.

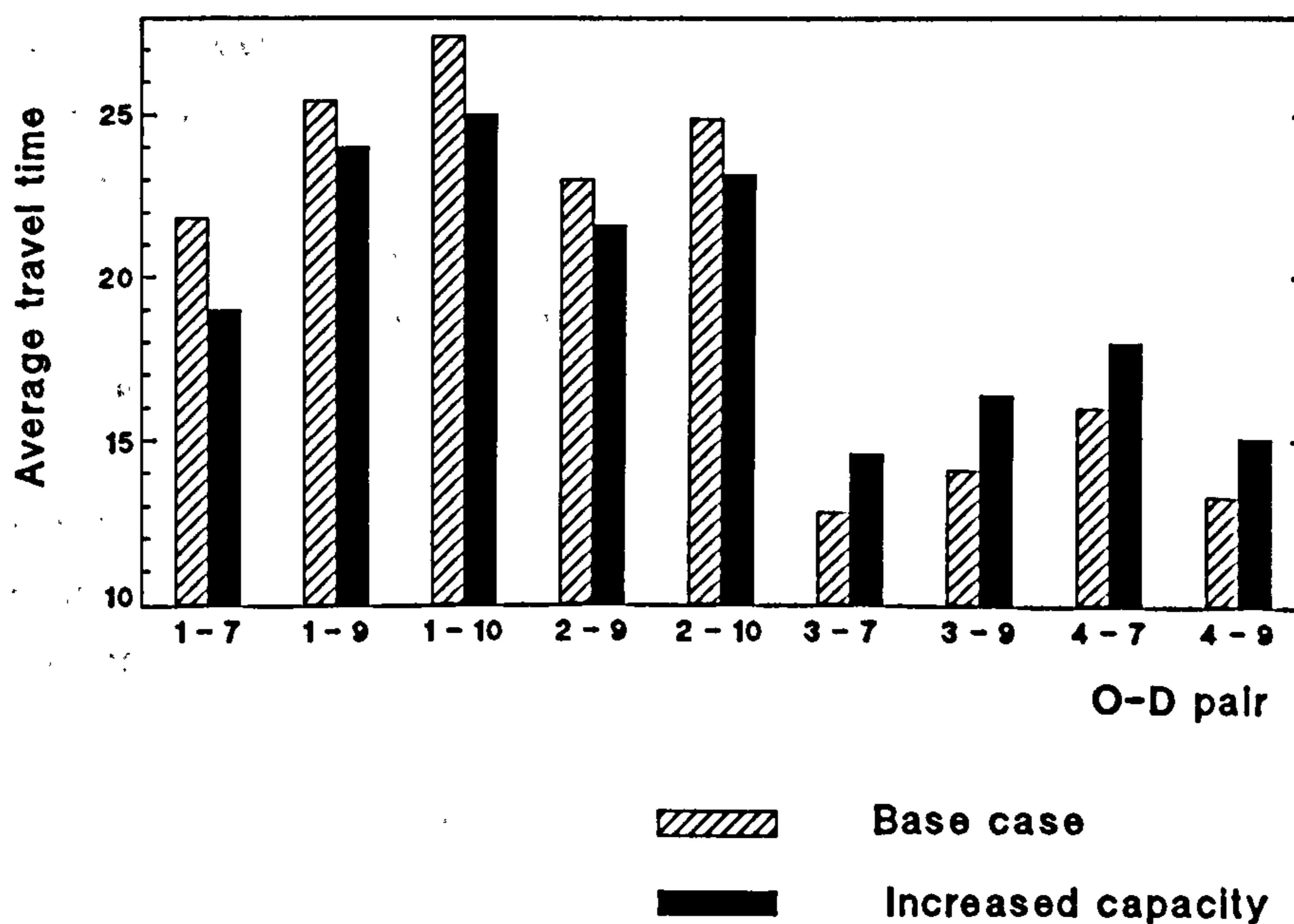


Figure 8.20 : Average O-D travel times before and after increasing the capacity of bottleneck 3.

### 8.7 Comparison of the demand adjustment and the equivalent formulation approach.

This section compares the outputs and the performance of the two approaches which can be used to solve the dynamic assignment problem. The first, described in chapter 6, defines the DSUE patterns using the demand adjustment mechanism (DAM) and will be termed the DAM approach; the second, described in chapter 7, uses the method of successive averages, and will be termed the MSA approach. The similarity between the DAM and MSA outputs was examined using the base case experiment; the correlation coefficient between the DAM and the MSA total (for the entire peak period) link flows was found to be 0.997. The similarity between the average link travel times derived from the DAM and the MSA approach was also investigated and the correlation coefficient has taken the same value. Furthermore, the correlation coefficient between the DAM and the MSA total path flows ( $=\sum_t Q_{k,rs}(t) \forall k,rs$ ) was found to be 0.985. Figure 8.21 demonstrates the similarity of the demand patterns derived from the DAM and MSA approach; the figure shows that the DAM and MSA time dependent departure rate distributions associated with the paths connecting the O-D pair 4-20 are identical. To study the rate of convergence of the DAM and the MSA approach, the maximum relative variation, RR (6.72), of the departure rate from one iteration to the next is examined versus the number of iterations performed. The effect of the number of iterations on the value of RR is illustrated in figure 8.22. For the first 15 iterations the MSA seems to behave better than the DAM. Thus for near optimal solutions, i.e., when the maximum acceptable variation of RR is relatively high, MSA may need fewer iterations than the DAM in order to converge. However if a higher degree of accuracy is desired, then the DAM seems to provide a better algorithm, since as it is shown in figure 8.22, DAM displays a more uniform rate of convergence than the MSA which is characterised by fluctuations of considerable magnitude. Therefore DAM seems to be a more efficient method in terms of the number of iterations required for an accurate solution. However, this approach requires path enumeration, which makes its use unfeasible for large scale networks. On the other hand, the MSA algorithm requires more iterations in order to converge to an equilibrium when the desired degree of accuracy is relatively high. Yet, it provides a framework which if used in combination with a DSNL mechanism which does not involve path enumeration will enable the application of dynamic assignment models to realistic networks.

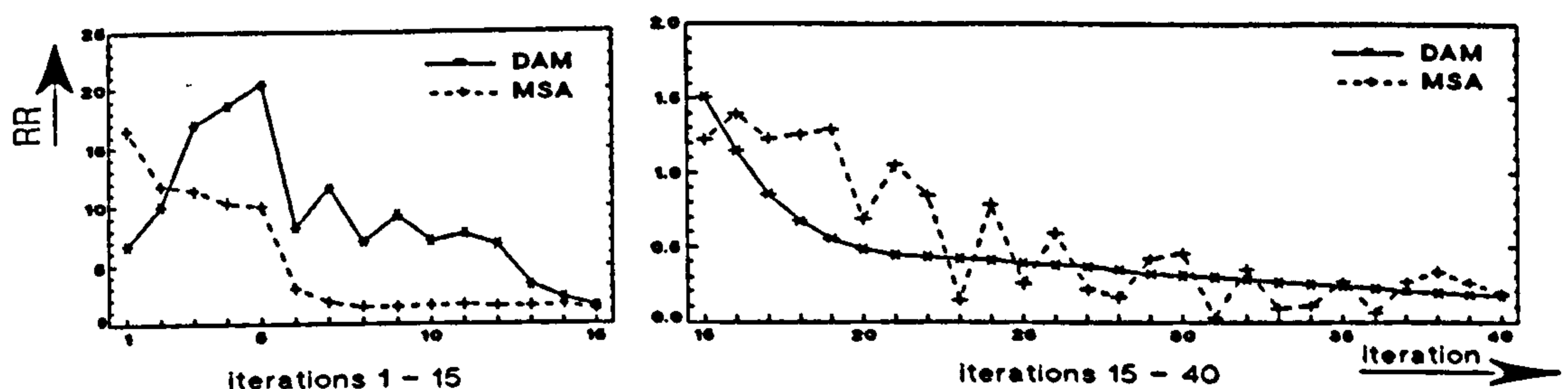


Figure 8.22: Convergence rate for the MSA and the DAM approach

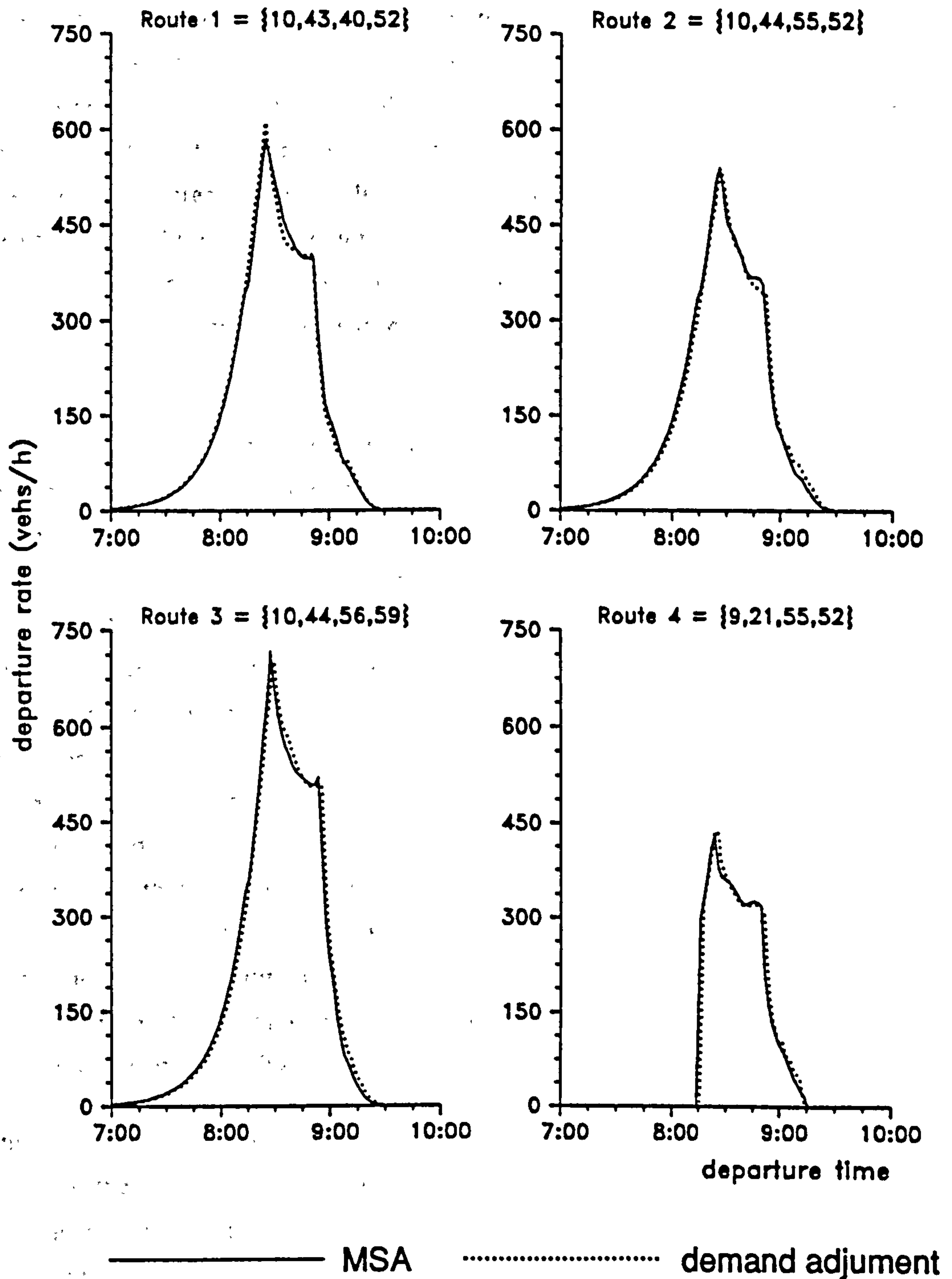


Figure 8.21: Comparison of departure rate distributions derived from the MSA and the demand adjustment mechanism approach.

### 8.8 Summary

In this chapter a number of simulation experiments were carried out to analyse the DSUE model outputs. The impact of changes in the work start time flexibility on traffic patterns and, the sensitivity of the model to changes in the level of demand and in the variability of travellers' preferences with respect to trip choices was investigated. The effects of increasing a bottleneck capacity on the evolution of the congestion patterns were analysed and a comparison of the two alternative methods to derive the equilibrium patterns, namely the one based on the demand adjustment mechanism and the one based on the method of successive averages was also presented.

The experiments have demonstrated the capability of the DSUE model to replicate the time varying traffic conditions during the morning peak, as opposed to the static assignment procedures which assume uniform traffic patterns and constant travel times. To demonstrate the model outputs the study was focused on the analysis of the travel time and departure rate time-varying distributions associated with various routes connecting the O-D pairs. For early departures the total disutility of travel is continuously decreasing, since the rate of increase of the travel time disutility is lower than the rate of decrease of the early schedule delay disutility; the departure rate distribution associated with a certain route is increasing and at the earliest departure time that imply on time arrivals at work, attains its maximum. After that time the departure rate distribution starts to continuously decrease until the departure time that is associated with the longest travel time. Then, as travel time starts decreasing, the departure rate distribution follow an increasing pattern and exhibit a second but lower peak at the latest departure time that results in on-time arrivals. After that time the rate of decrease of the travel time disutility is lower than the rate of increase of the late schedule delay disutility and thus departure rates start to decrease again.

Changes in the flexibility of the work start time are shown to have an important impact on the level of congestion developed in the network. The major effect of increasing the work start time flexibility is to spread the departure rate distribution over a longer time period, and thus to reduce the level of the traffic congestion and O-D travel times. Under inflexible working time schedules, the departure rate distributions exhibit a single and much higher peak at the departure time that results in an on-time arrival. As a result, the levels of congestion are increased due to the high concentration of traffic within a short time interval.

As expected, increases in the overall level of demand cause severe congestion problems on all the already congested routes. The shape of the departure rate and travel time distributions remain the same but the distributions are more highly peaked. Furthermore, as the level of demand increases the mode of the departure rate distribution shifts to the left, revealing evidence that travellers respond to the higher level of congestion, resulting from the increased level of demand and the limited network capacity, by shifting towards earlier departure times. Peak period travel times are substantially higher and the congestion period is extended over a longer period towards earlier departures.

Increased variability of preferences with respect to route and departure time choice result in lower levels of congestion and lower travel times but higher levels of schedule delay; this reflects the tendency of a higher proportion of travellers to select "non optimum" routes and departure times. Within an environment where travellers have a better perception of the utility associated with each combination of route and departure time choice, and act as more rational decision makers, the level of congestion increases. This, since they want to maximise their utility of travel, and therefore their choices are concentrated in the vicinity of the highest utility alternatives. Thus, departure rate distributions are less spread and exhibit high peaks, and the highest proportion of the demand is concentrated on shortest routes which therefore get congested. However, although travellers with lower variability of preferences experience longer travel times, on the whole they are better off since their trip choices result in low levels of schedule delay which in turn imply higher levels of the total utility of travel.

Simulation experiments using the queueing theory based travel time model have demonstrated the ability of the DSUE model to represent the evolution of queues at different locations of the network during varying time intervals. The model represents the time dimension of traffic and thus in contrast to static assignment procedures it takes into account the fact that cars which are held at a bottleneck cannot at the same time cause congestion downstream. The experiments have shown that increasing a bottleneck capacity may result in a shift of the demand towards later departures and may cause congestion in downstream links due to the traffic that is no more held in the expanded capacity bottleneck. As a result O-D travel times for some O-D pairs may decrease, while for others they may increase.



# 9 using the DSUE model for policy analysis

### **Objective**

The purpose of this chapter is to demonstrate the potential applications of the DSUE model as a policy analysis tool; the policy measures examined include road pricing and route guidance.

### **9.1 Introduction**

Over the last decades a number of approaches such as building more roads, creating high occupancy vehicle lanes, and promoting car pooling and public transportation have been used as alternative measures to reduce traffic congestion and the pollution it generates. However in most cases these approaches have not achieved substantial improvements. Under these conditions, it was considered that other forms of transportation systems management should also be pursued.

Thus more recently, traffic restraint measures and applications of information technology have been seen to offer potential for combating congestion. These approaches have the advantage of not creating any conflicts with other traffic management strategies and can also be used in conjunction with them.

Various types of traffic restraint measures are identified and reviewed by Jones (1989). In a number of these measures a charge on a link or area basis is applied for the use of roads, with the aim to discourage travellers from using those parts of the network that are severely congested. On the other hand, route guidance systems aim to achieve a better utilisation of the existing facilities by enhancing drivers' information on the traffic conditions they will encounter during the course of their journey and therefore enabling them to select optimum routes. Both systems require advanced technology

hardware and therefore are associated with high investments which make their application in small networks rather unfeasible.

Furthermore, up to date, the majority of the research effort in the area of road pricing and route guidance has been concentrated on the design of the hardware and software required for their operation. However the effectiveness, or in other words the degree to which these systems will improve existing traffic conditions has not been analysed thoroughly.

Thus, the uncertainty on the potential benefits derived from the implementation of such strategies on the one hand, and the high investments they require on the other, argue for the development of techniques that will estimate their impact on travel patterns and the overall level of congestion in urban road networks.

This chapter aims to demonstrate the potential applications of the DSUE model as a tool useful to forecast the effects of road pricing and route guidance. The analysis is based on the results of simulation based experiments where a variety of policies is examined and their effects on the peak travel demand patterns are analysed.

Thus, in the following section the DSUE assignment model is used to analyse the effects of alternative road pricing policies. In section 9.3 a method for using the DSUE model to evaluate the effects of a route guidance system is discussed and finally section 9.4 concludes the chapter.

## **9.2 Evaluating the effects of road pricing**

In this section the effects of a variety of road pricing policies are explored with the intention to indicate the sensitivity of route and departure time decisions. The purpose is to derive user equilibrium on a simple network and to evaluate the efficiency gains from various types of tolls.

Pretty (1988) points out that the possible reactions of an existing traveller to road pricing are:

- to make the trip and pay,
- to not make the trip,
- to make the trip by a different route,
- to change the time of travel,
- to change origin or destination,
- to form a car pool,
- to change mode away from private car travel.

In the analysis presented in this section a traveller's response to the road pricing scheme is limited to the choice of an alternative route (not requiring toll) or a shift to a departure time (avoid toll). In other words the problem is simplified by considering that trip origins and destinations are fixed and the total demand for travel during the peak is inelastic (i.e. no switching to alternative mode). However the model can be easily extended to treat this dimension of choice. Such an extension of the model is given in Vythoukias (1990b), where the DSUE was developed to evaluate the effects of road pricing as a part of the research carried out within the project Integrated Demand Management Strategies (V1008) sponsored by the DRIVE programme of the Commission of European Communities.

Road pricing schemes can be classified in the following categories (Jones, 1989):

- Toll charges for using a section of road ( or crossing a cordon line).
- An area licence fee, giving the right to use a vehicle in an otherwise prohibited area.
- A road pricing charge - either a sophisticated version of the above, or a system based on mileage or time spent in the area.

The analysis presented below focuses on the examination of the effects of two road pricing schemes both falling in the first category. Thus in the first set of experiments toll charges are implemented for using specific links of the network, and in the second a toll is charged when crossing a cordon line, or more specifically when entering the toll ring. The links selected for the first set of experiments are the most congested ones; these are links 8, 17, 40, 44, 48, 51, 52, 54, 55, 56 and 62 . The toll ring, illustrated in figure 9.1, covers the area around nodes 17 and 18, which is one of the most congested areas of the network; vehicles are charged for inbound journeys, and thus a toll is charged for using links 40, 44, 55, 58, 62 and 48. In the experiments carried out the toll was considered to be constant as well as to vary over the course of the peak. Details of the different "links" and "ring" road pricing scenarios used in the experiments are shown in table 9.1. Thus for example scenario TR115 implies that a toll of £ 1.00 is charged when a driver enters the toll ring during the interval 8:00 to 8:30, and a charge of £ 1.50 if he enters the ring during the interval 8:30 to 9:30.

In order to evaluate the effects of the different road pricing scenarios, the utility function used in the DSUE model was modified to incorporate a new term reflecting the disutility due to the toll which is assumed to increase linearly at a rate of  $\kappa$  . The new utility function has the form :

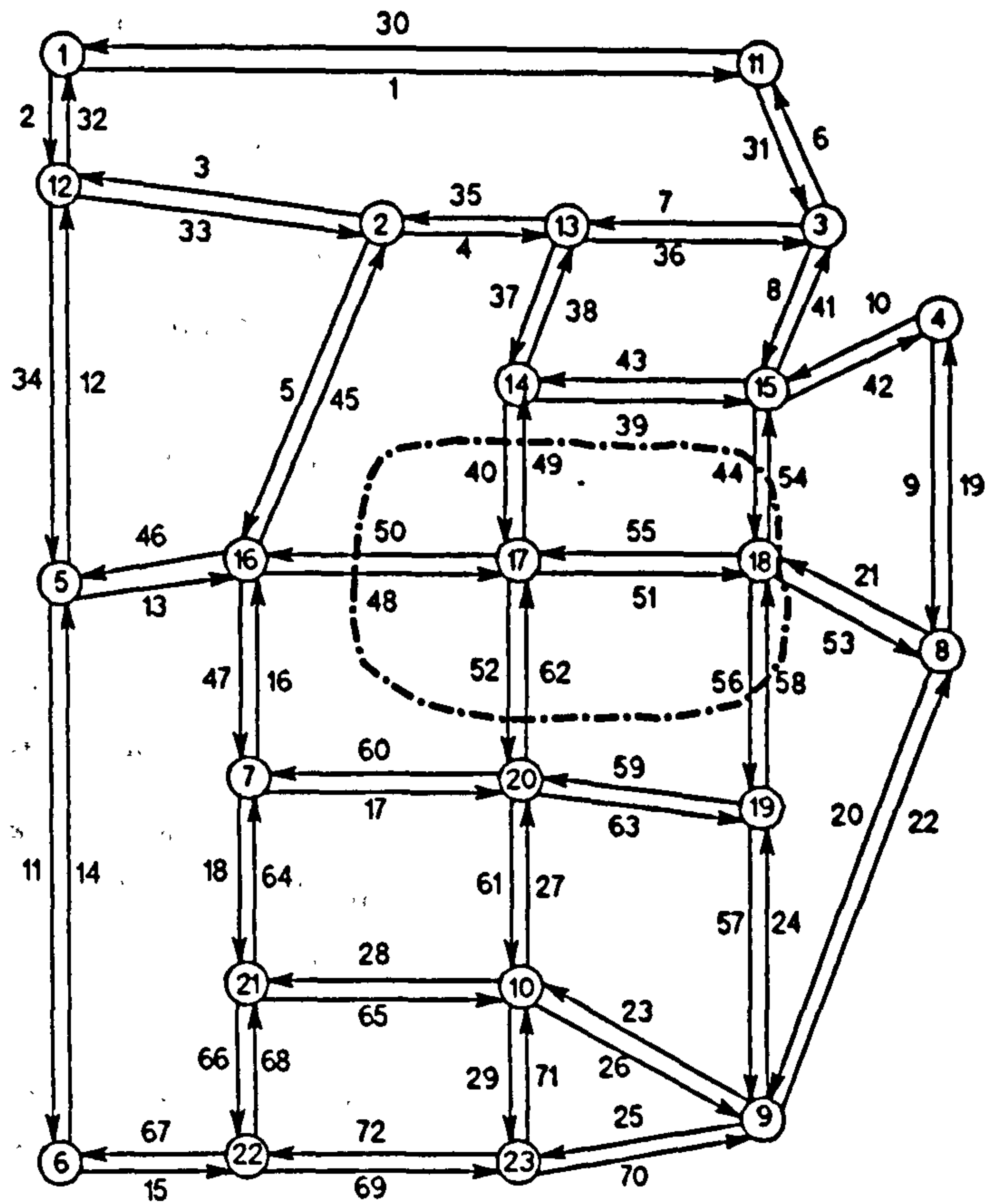


Figure 9.1 : The toll ring

scenario	area	time period	
		8:00-8:30	8:30-9:30
TR010	ring	-	1.0
TR015	ring	-	1.5
TR020	ring	-	2.0
TR110	ring	1.0	1.0
TR115	ring	1.0	1.5
TR120	ring	1.0	2.0
TL010	links	-	1.0
TL015	links	-	1.5
TL020	links	-	2.0
TL110	links	1.0	1.0
TL115	links	1.0	1.5
TL120	links	1.0	2.0

Table 9.1 :Toll charges in pounds for the alternative road pricing scenarios

$$VT_{k,rs}(t) = V_{k,rs}(t) - \kappa \cdot TL_{k,rs}(t) \tag{9.1}$$

where  $V_{k,rs}(t)$  is the utility expressed by equation (6.55) and  $TL_{k,rs}(t)$  is expressed as:

$$TL_{k,rs}(t) = \begin{cases} \sum_{j \in K_{rs}(t)} tl_j(t) & \text{if the route } k \text{ includes a link } j \text{ and departing from } r \text{ at} \\ & \text{time } t \text{ implies that the driver will enter link } j \text{ at time } t'. \\ 0 & \text{otherwise} \end{cases}$$

where  $tl_j(t)$  is the toll charged when entering link  $j$  at time  $t$ .

To illustrate the model outputs the O-D pair 4-20 is used as an example; the equilibrium distributions of departure rates and travel times under the scenarios TR010 and TR020 are presented in figures 9.2 and 9.3. The figures in conjunction with tables 9.2a and 9.2b illustrate the response of travel demand to the two road pricing policies.

Table 9.2a: Demand and attributes of routes connecting the O-D pair 4-20, under scenario TR010

O/D PAIR : 4 - 20		Total Demand = 2800.					
Path	Links	Demand	Av. Trav. Time	FreeFlow Time	Av. Sch. early	Delay late	Average Dis.x100
1	10 - 43 - 40 - 52 -	342.	15.9	12.0	12.9	0.4	323.5
2	10 - 44 - 55 - 52 -	340.	16.2	12.7	12.2	0.4	324.0
3	10 - 44 - 56 - 59 -	397.	15.6	13.3	11.3	0.3	300.7
4	9 - 21 - 55 - 52 -	144.	20.6	16.1	1.7	0.6	345.5
5	9 - 21 - 56 - 59 -	479.	19.8	16.7	1.1	0.6	233.6
6	9 - 20 - 24 - 59 -	320.	21.1	18.9	0.0	0.7	243.1
7	9 - 20 - 25 - 71 - 27 -	364.	20.4	19.1	0.0	0.6	233.0
8	9 - 20 - 23 - 27 -	413.	19.5	18.3	0.0	0.6	222.6
Total		2800.					
Average			18.5	15.9	4.9	0.5	270.3

Time Interval	Paths
0 - 75	1, 2, 3
75 - 85	1, 2, 3, 4, 5,
85 - 135	1, 2, 3, 4, 5, 6, 7, 8
135 - 145	1, 2, 3, 4, 5

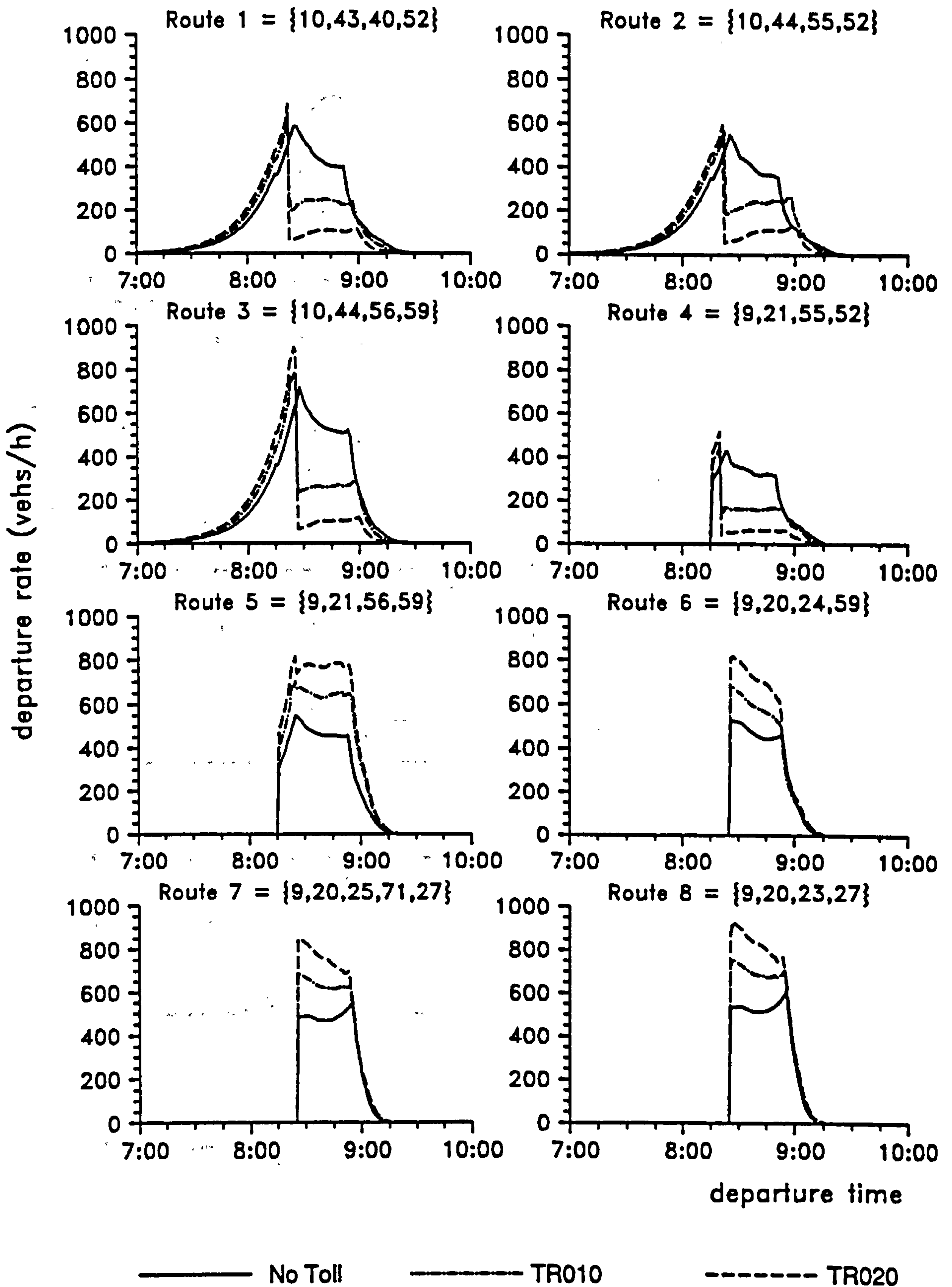


Figure 9.2: Departure rate distributions for O-D pair 4-20 under the no-toll, TR010 and TR020 scenarios.

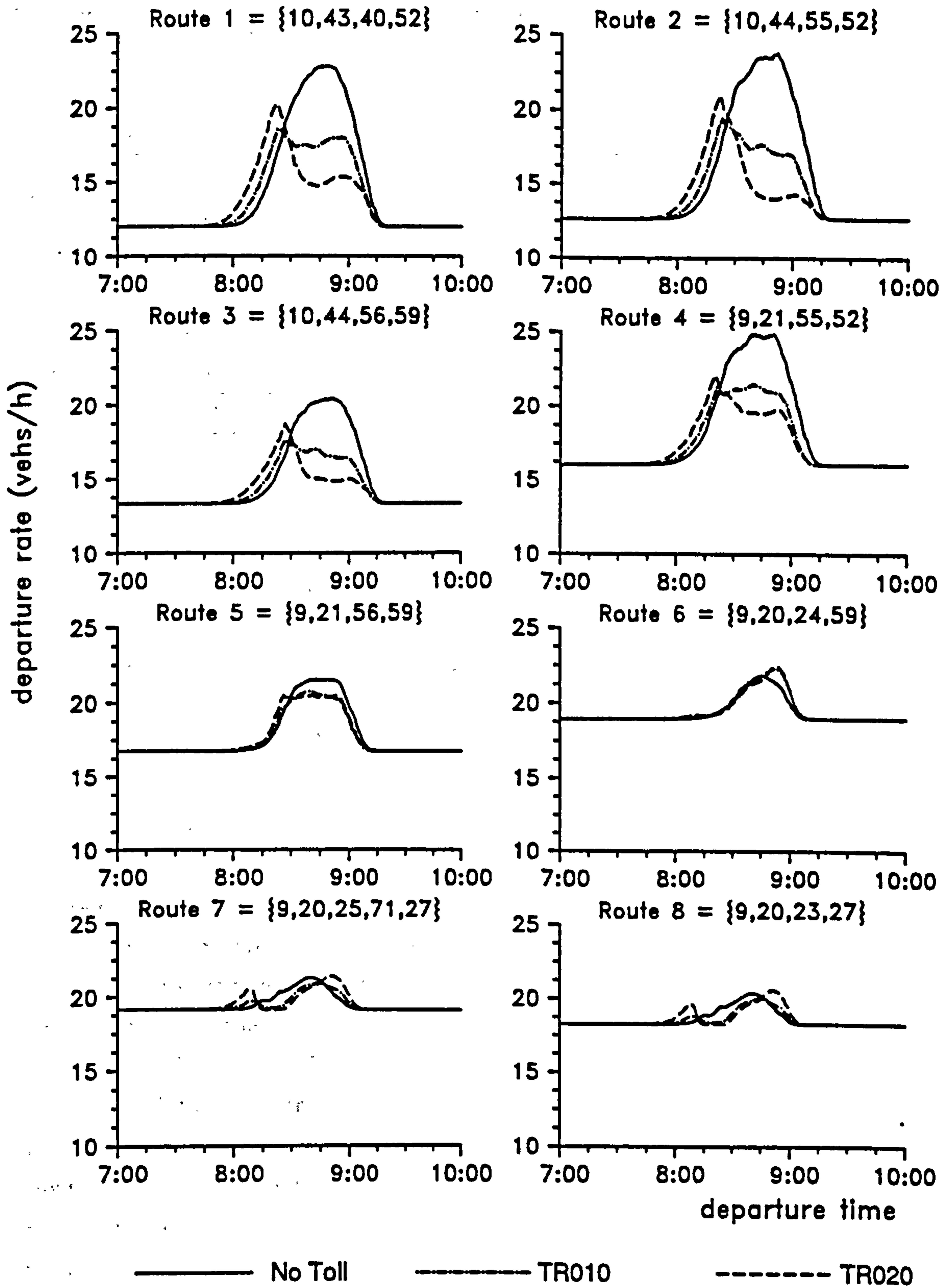


Figure 9.3: Travel time distributions for O-D pair 4-20 under the no-toll, TR010 and TR020 scenarios.



**Table 9.2b: Demand and attributes of routes connecting the O-D pair 4-20, under scenario TR020**

O/D PAIR : 4 - 20		Total Demand = 2800.					
Path	Links	Demand	Av. Trav. Time	FreeFlow Time	Av. Sch. early	Delay late	Average Dis.x100
1	10 - 43 - 40 - 52 -	272.	15.7	12.0	17.4	0.2	349.4
2	10 - 44 - 55 - 52 -	261.	15.8	12.7	17.1	0.2	356.8
3	10 - 44 - 56 - 59 -	333.	15.6	13.3	15.1	0.1	316.7
4	9 - 21 - 55 - 52 -	80.	20.5	16.1	2.7	0.4	379.5
5	9 - 21 - 56 - 59 -	566.	19.9	16.7	1.0	0.6	234.3
6	9 - 20 - 24 - 59 -	388.	21.0	18.9	0.0	0.7	242.0
7	9 - 20 - 25 - 71 - 27 -	422.	20.5	19.1	0.0	0.6	233.5
8	9 - 20 - 23 - 27 -	479.	19.6	18.3	0.0	0.6	223.6
Total		2800.					
Average			18.8	16.4	5.4	0.5	270.0

Time Interval	Paths
0 - 75	1, 2, 3
75 - 85	1, 2, 3, 4, 5
85 - 130	1, 2, 3, 4, 5, 6, 7, 8
130 - 135	1, 2, 3, 4, 5, 6
135 - 145	1, 2, 3

A major shift of demand is observed from the shortest routes 1,2 and 3, (which however cross the cordon line) towards the longest no-toll routes 5,6,7 and 8. The former set of routes also demonstrates the shifting peaks phenomenon, i.e. the potential development of a congestion period just before the toll period. Drivers departing before 8:20 and travelling from node 4 to node 20 use routes 1, 2 and 3 which are associated with the lower travel times. Peaks along these routes are much higher than in the no toll case since the majority of drivers attempt to enter the ring before 8:30. Departing after 8:20 would imply a toll charge and therefore after this departure time there is an abrupt reduction in the demand for using routes that cross the cordon line and a high increase in the demand for longer routes which however do not enter the toll ring. A similar behaviour in the departure rate and travel time distributions is observed (figures 9.4 and 9.5) under the scenarios TR110, and TR120. Here the shift to the longer routes 5,6,7 and 8 is even higher and drivers following routes 1,2 and 3 are more evenly distributed over the morning commute period. Departure rate distributions

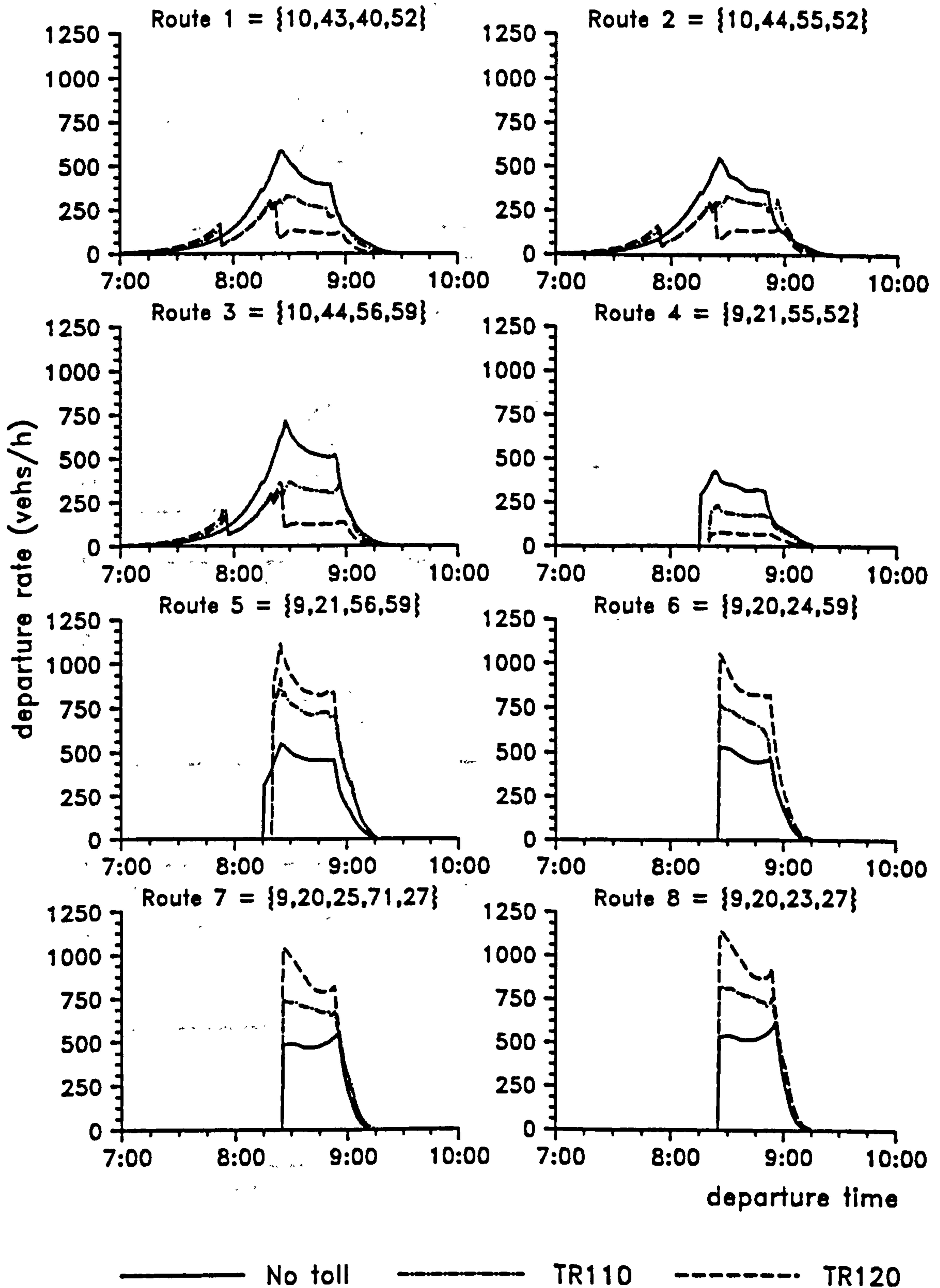


Figure 9.4: Departure rate distributions for O-D pair 4-20 under the no-toll, TR110 and TR120 scenarios.

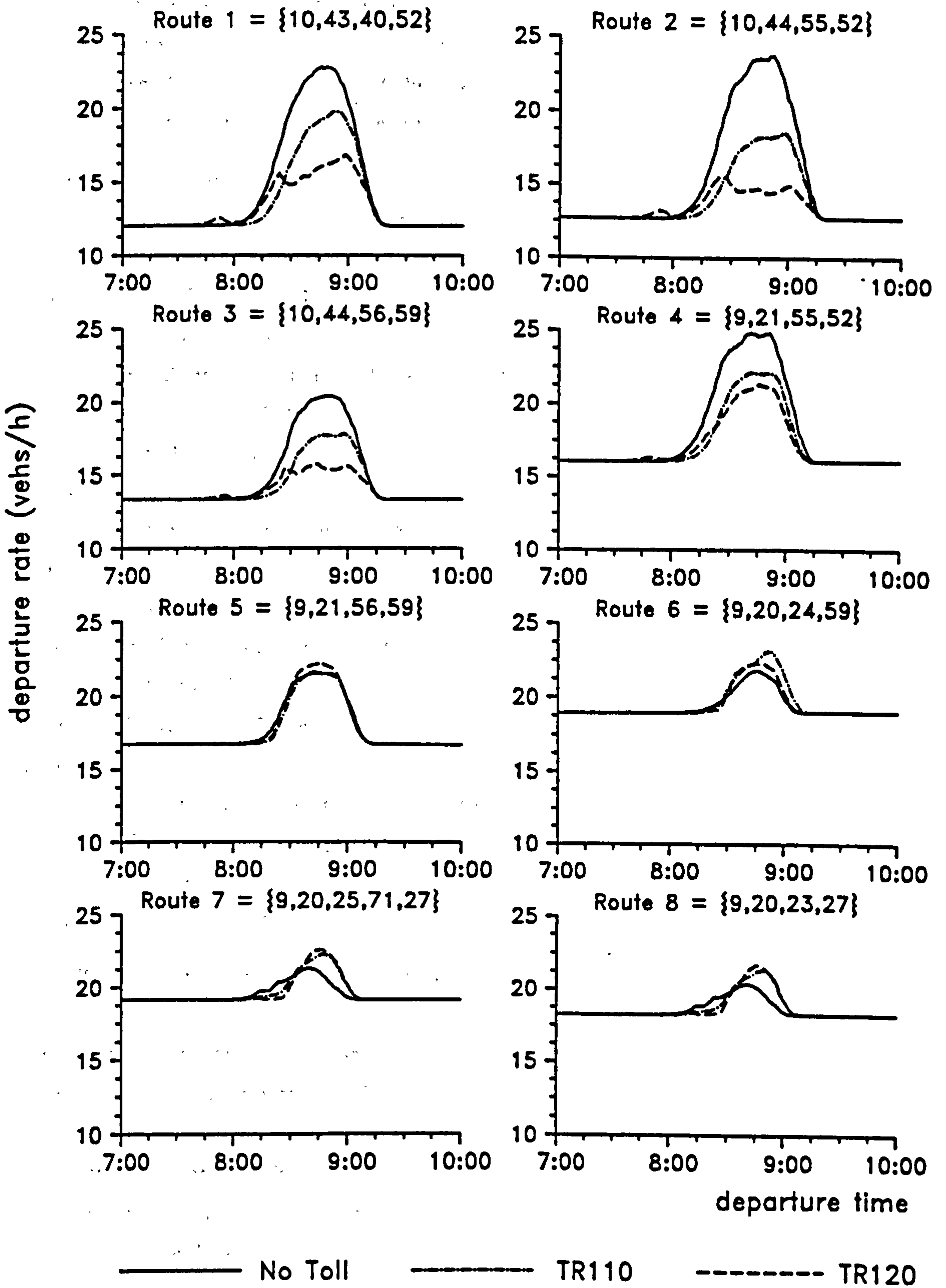


Figure 9.5: Travel time distributions for O-D pair 4-20 under the no-toll, TR110 and TR120 scenarios.

for the latter routes exhibit two peaks: one before 8:00 and one before 8:30 due to the variation in the toll charge. The later peak (at about 8:20) is higher, although a toll of £ 1.0 is charged during the period before this peak, since departures before 8:00, although not incurring a toll charge, result in very high early schedule delays.

The abrupt discontinuities in the departure rate distributions can also be explained using figure 9.6. This figure illustrates the departure time dependent travel times, departure rate distributions, disutilities and schedule delays associated with the route 1 connecting the O-D pair 4-20 under the scenarios TR010 and TR115. Initially the disutility is decreasing since later departures do not imply a rapid increase in travel times, but result in a substantial decrease of the total disutility due to the reduction in early schedule delay. Thus, since the rate of decrease of the early schedule delay disutility is higher than the rate of increase of the travel time disutility, departure rates are continuously increasing up to the departure time which results in crossing the cordon line during the toll period. At this departure time an increase in the disutility level is observed which is attributed to the additional disutility due to the toll charge, and causes an abrupt discontinuity in the departure rate distribution. The number of drivers departing per minute is substantially reduced and remains at an almost constant low level until the departure time that implies late arrivals. After this departure time the rate of increase of the late schedule delay disutility is higher than the rate of decrease of the travel time disutility and thus departure rates start to continuously decrease. Similar shapes exhibit the distributions associated with the TR115 scenario, with the difference that two peaks are observed in the departure rate distribution due to the two discontinuities in the disutility distribution which are attributed to the two different level of tolls imposed during the intervals [8:00, 8:30] and [8:30, 9:30].

The effects of the different road pricing policies on the traffic conditions observed in the network links are illustrated in figure 9.7 by referring to the time dependent distribution of the number of cars and corresponding speed within links 40, 44, 48, 52, 55 and 62, under the no-toll, TR020 and TR120 scenarios. The figure shows that under a "short duration - high toll" strategy higher levels of congestion may develop just before the toll period. The observed high link density even after the start of the toll period is caused by the vehicles which have entered the link before the toll period but haven't exit the link due to the high congestion. The figures also demonstrate that congestion develops earlier and although it may attain higher levels, it has a shorter duration. A "longer duration - differentiated toll" scheme substantially reduces the level of congestion. Peaks in traffic patterns still occur but remain at a lower level. The congestion period extends almost over the same period as in the no-toll case; the

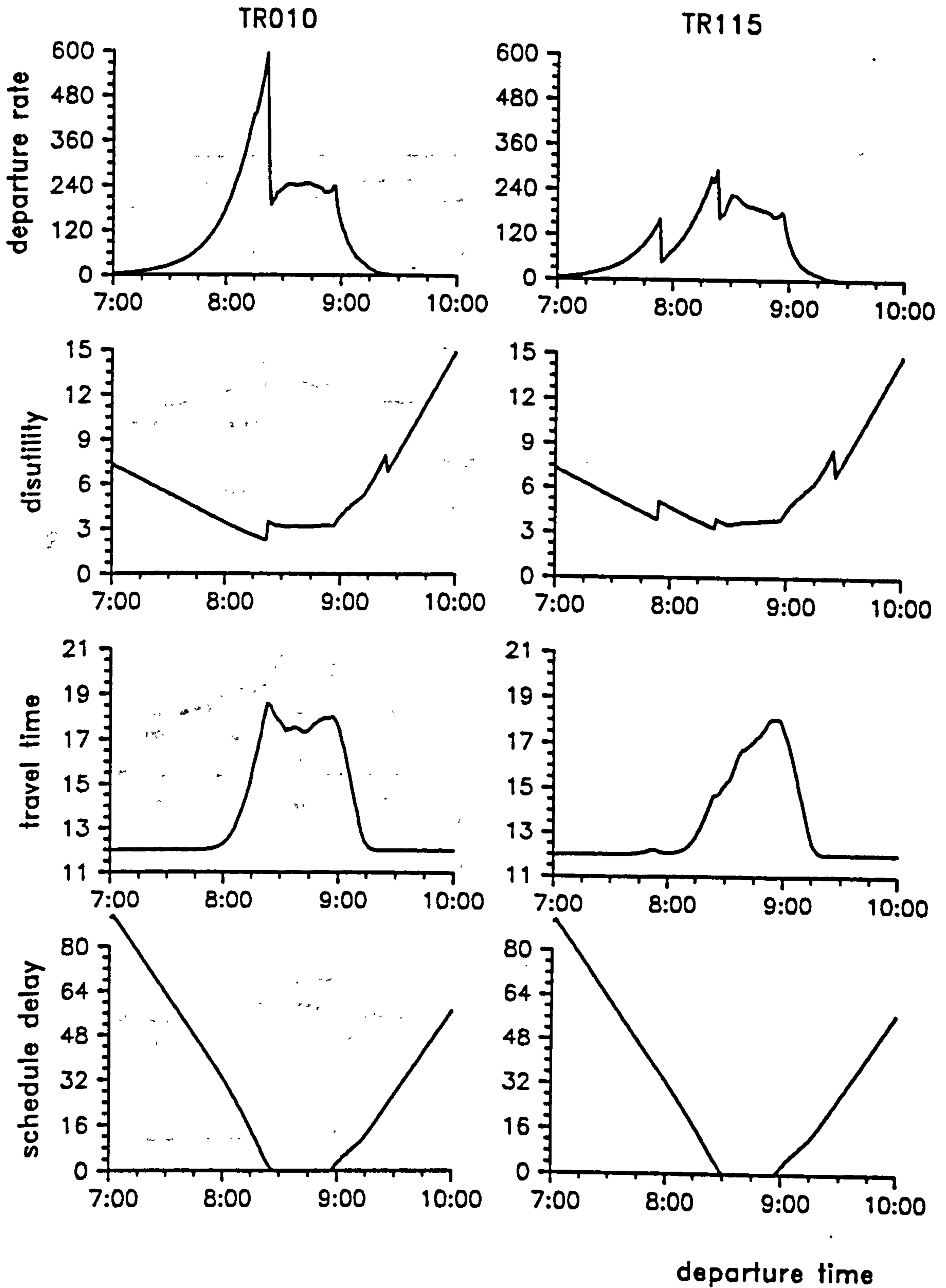


Figure 9.6: Departure rate, disutility, travel time and schedule delay distributions for route 1 connecting O-D pair 4-20 under scenarios TR010 and TR115.

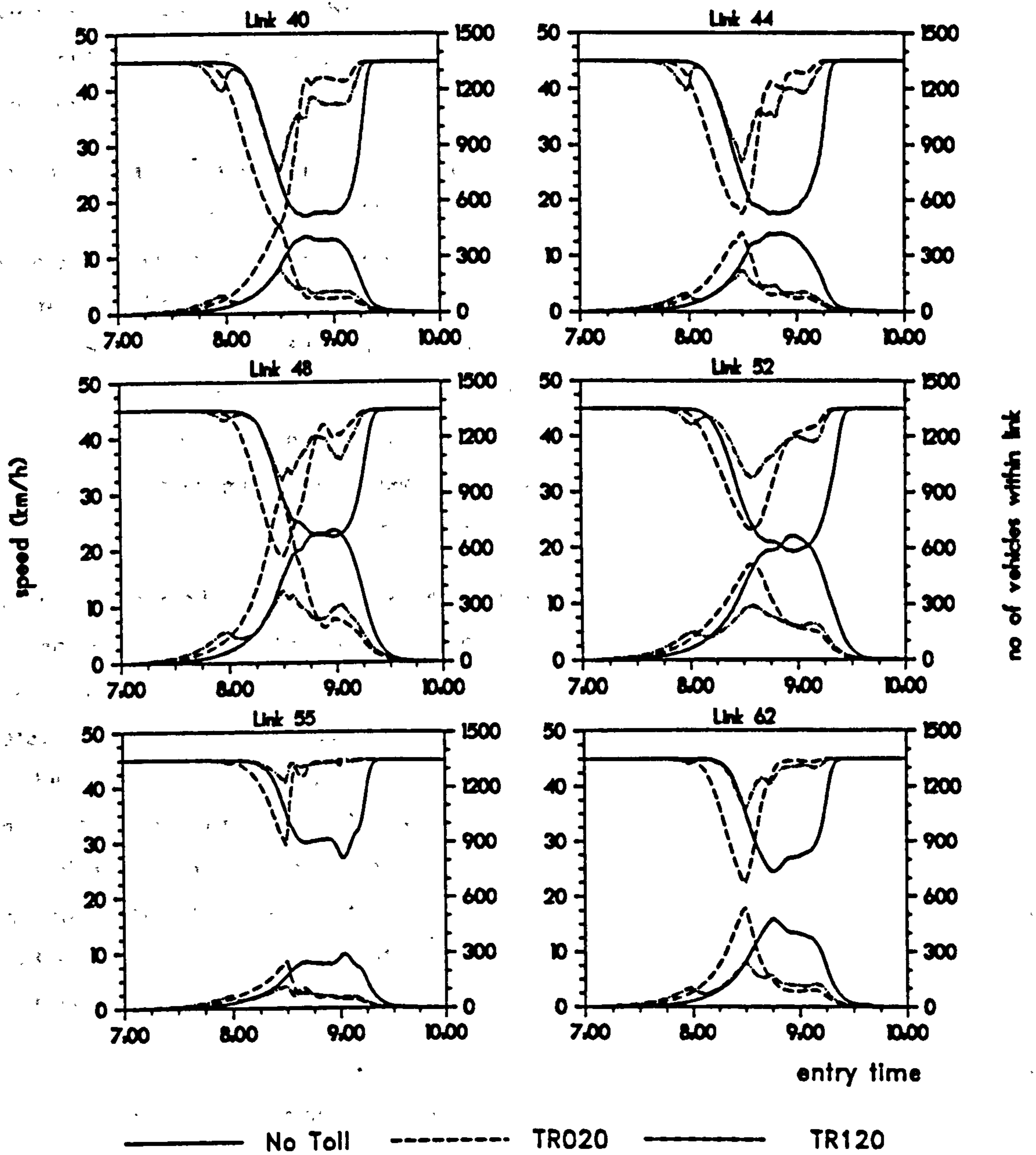


Figure 9.7: Time varying distributions of speeds and number of vehicles within various links of the network under the no toll, TR020 and TR120 scenarios..

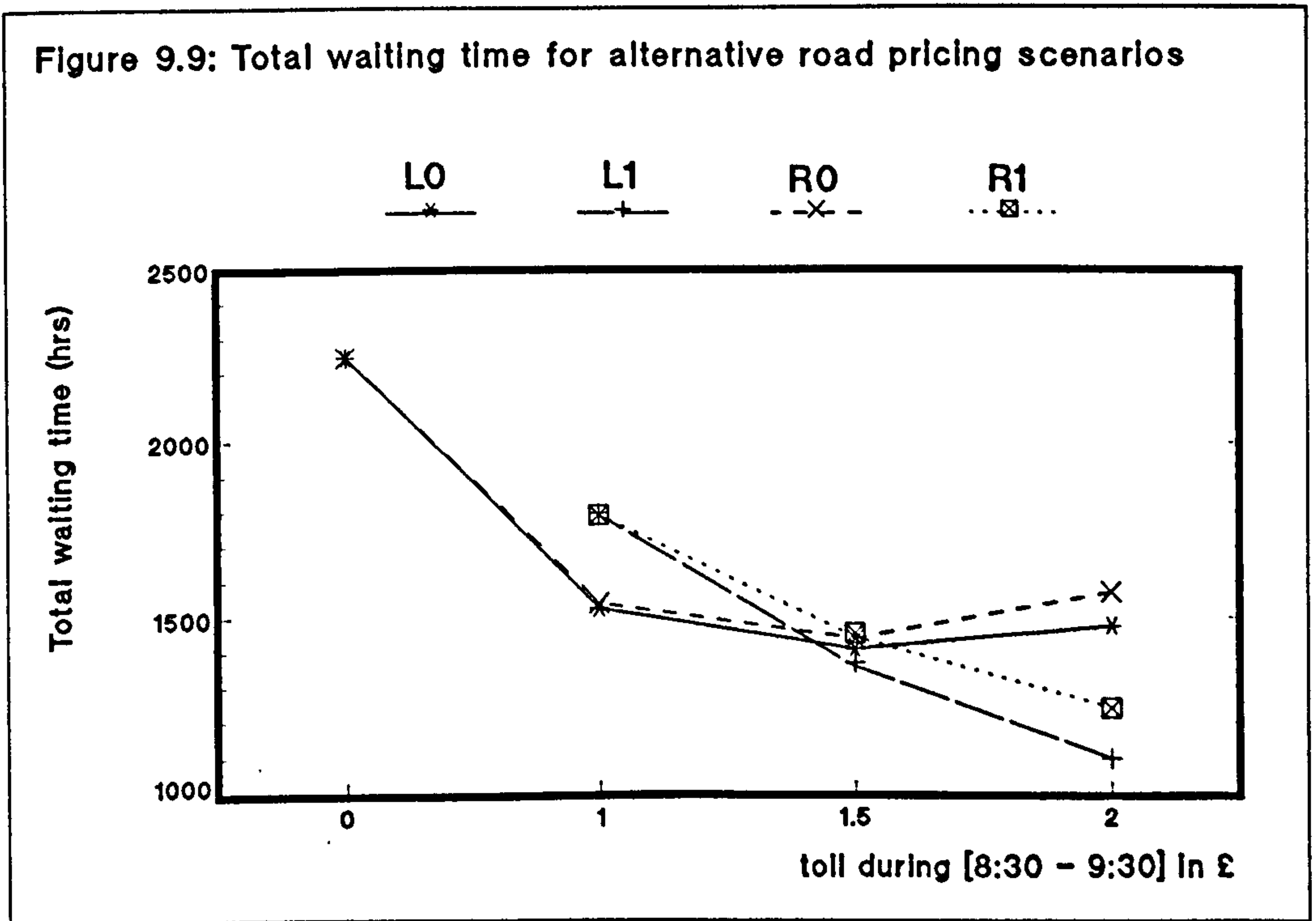
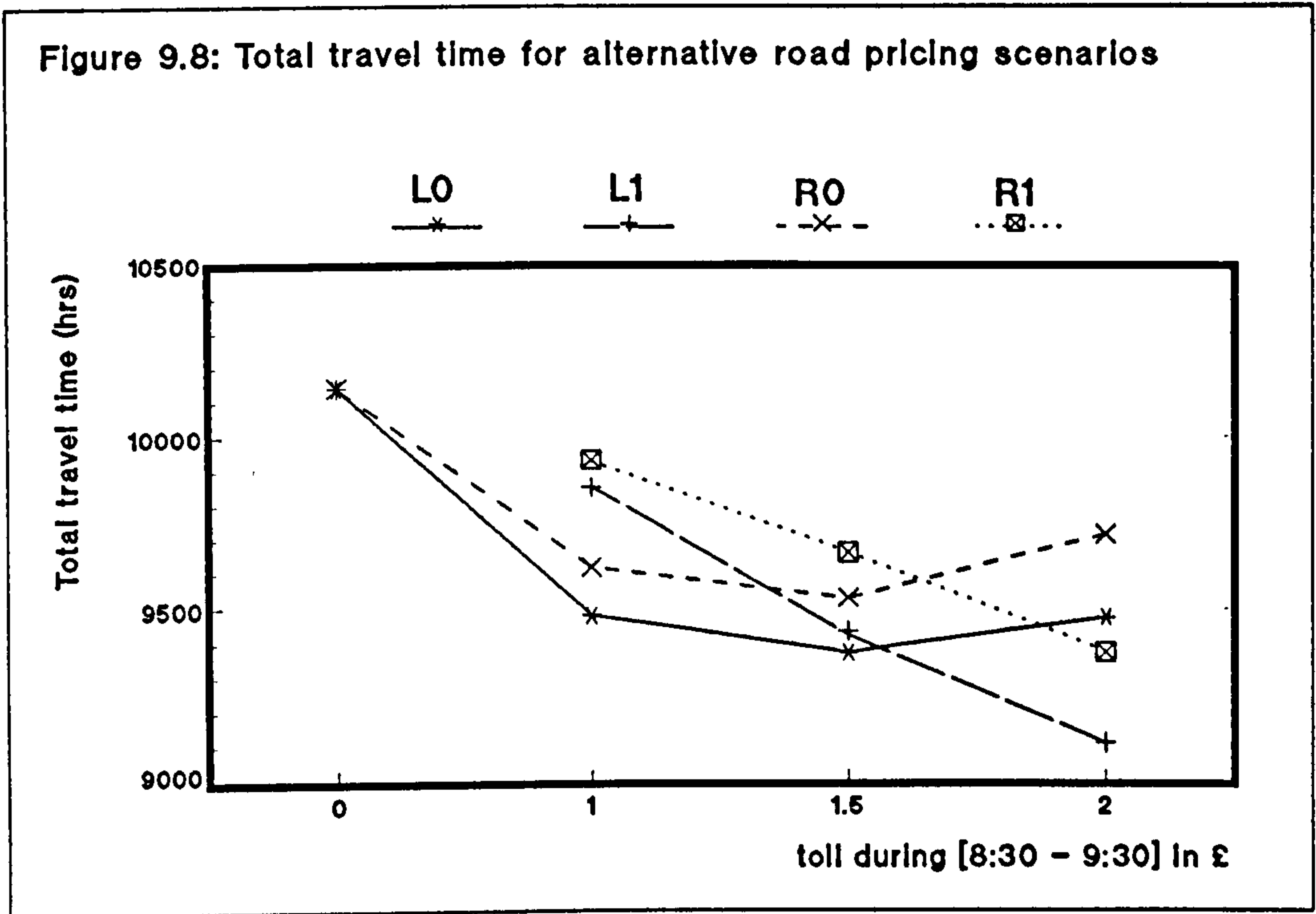
highest congestion levels occur during the "lower toll" time period while during the "higher toll" congestion is kept at very low levels.

In general, the various time dependant distributions describing demand, route and link characteristics which were obtained from the set of experiments dealing with the implementation of a toll charge for using selected links in the network, exhibit similar behaviour

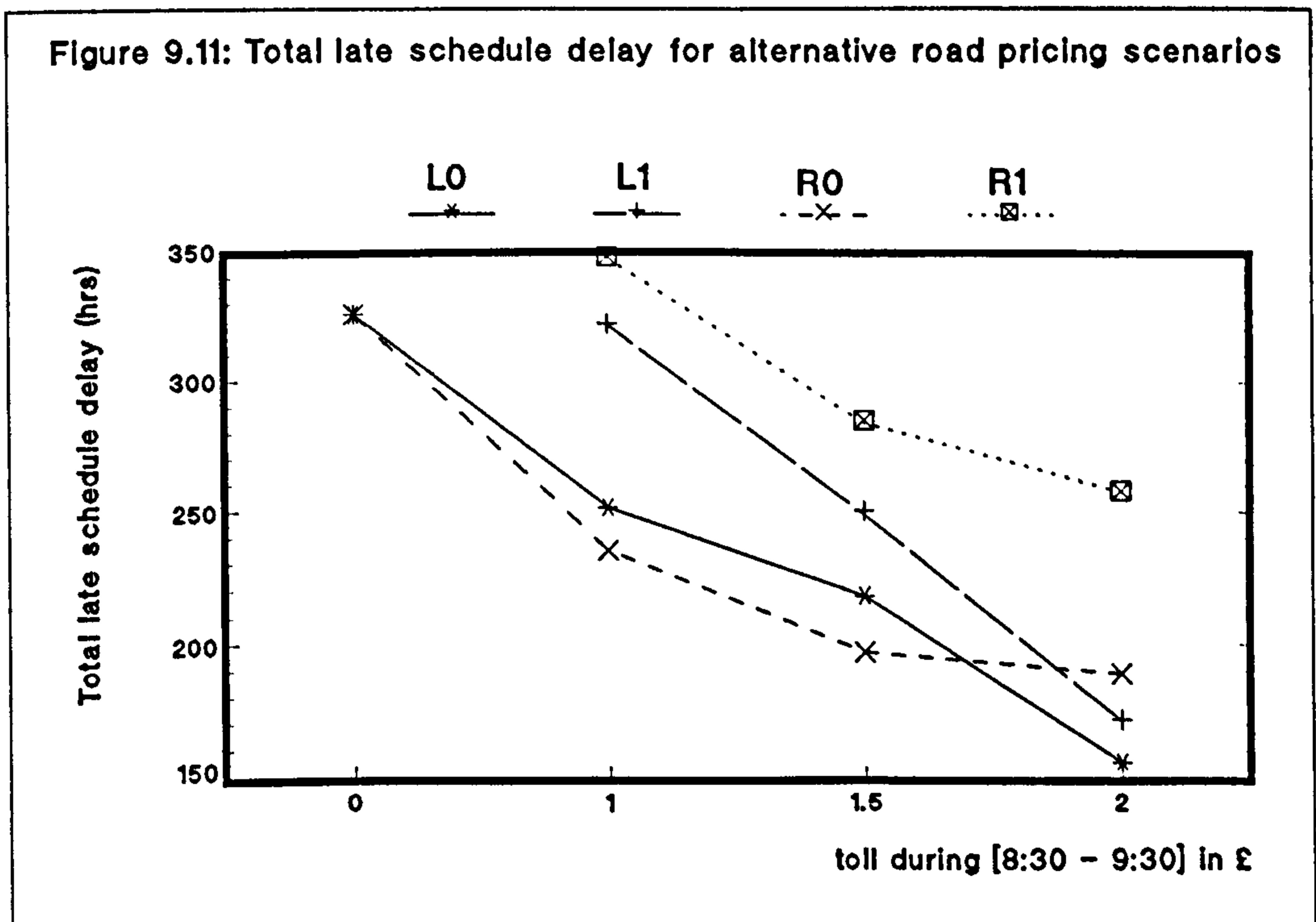
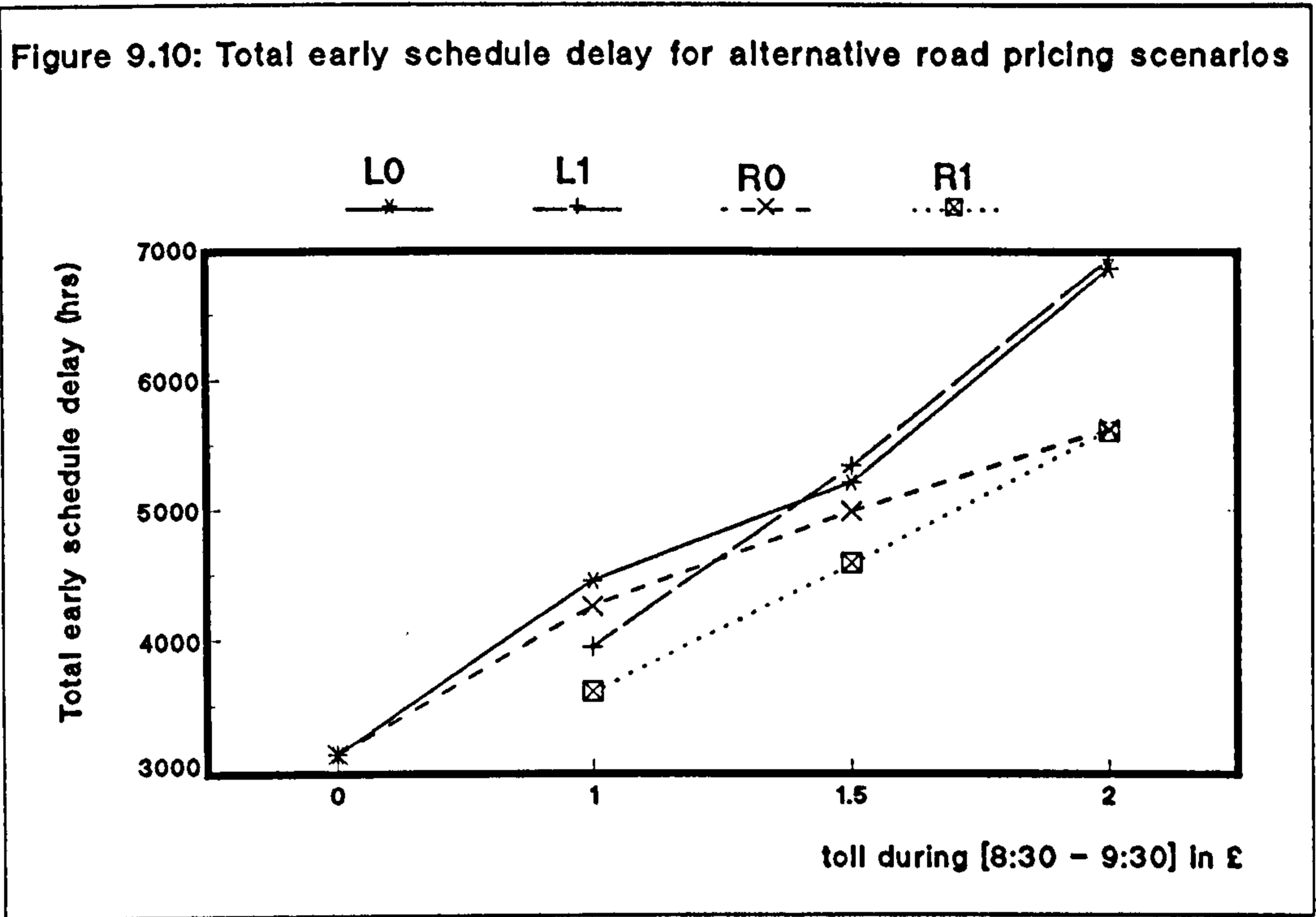
One consideration in performing this analysis is also to evaluate the performance of different road pricing policies. The performance is measured in terms of reduction in the total travel and waiting time, early and late schedule delay, and the revenue collected by the highway authority. The results are illustrated in figures 9.8, 9.9 . . . 9.12. In these figures the different scenarios are classified in the categories L0, L1, R0 and R1; in this notation the letter (L or R) represents a "links" or a "ring" strategy, and the number (0 or 1) represents the level of the toll charged during the period [8:00, 8:30]. Thus, for example, the category L0 includes the scenarios TL010, TL015 and TL020, and the category R1 the scenarios TR110, TR115 and TR120.

Under all the road pricing policies examined, the total travel time was substantially reduced from the 10140 hrs calculated in the no toll case. The figures illustrate that in general for the same policy in terms of period of charge and level of toll, a "links" strategy results in lower travel and waiting times than the corresponding "ring" strategy. On the other hand, the latter strategy results in lower early schedule delays when compared to the former. This because a "links" strategy affects a higher proportion of the total number of drivers (since the toll is charged in more links) who shift to earlier departures and alternative routes in order to avoid the toll. As a result of this more broad distribution of the demand in space and time, drivers experience higher early schedule delays and lower waiting times.

Figures 9.8 and 9.9 illustrate that for the policies that do not implement any toll during the period 8:00 to 8:30, travel and waiting times are not monotonic functions of the level of toll imposed in travellers during the period 8:30 to 9:30. In both the "ring" and the "links" scheme, the best performance in terms of savings in travel and waiting times is achieved for a £1.5 toll. For higher levels of charge the delays start increasing due to the shift of the peak period before the toll period. The higher level of congestion is attributable to the greater proportion of drivers who are willing to experience long early schedule delays than pay the £2.0 toll. Similar phenomena are not observed when the toll period extends over the period 8:00 to 9:30. This since a shift to departures which imply no toll charge (i.e., earlier than 8:00) would imply very long early

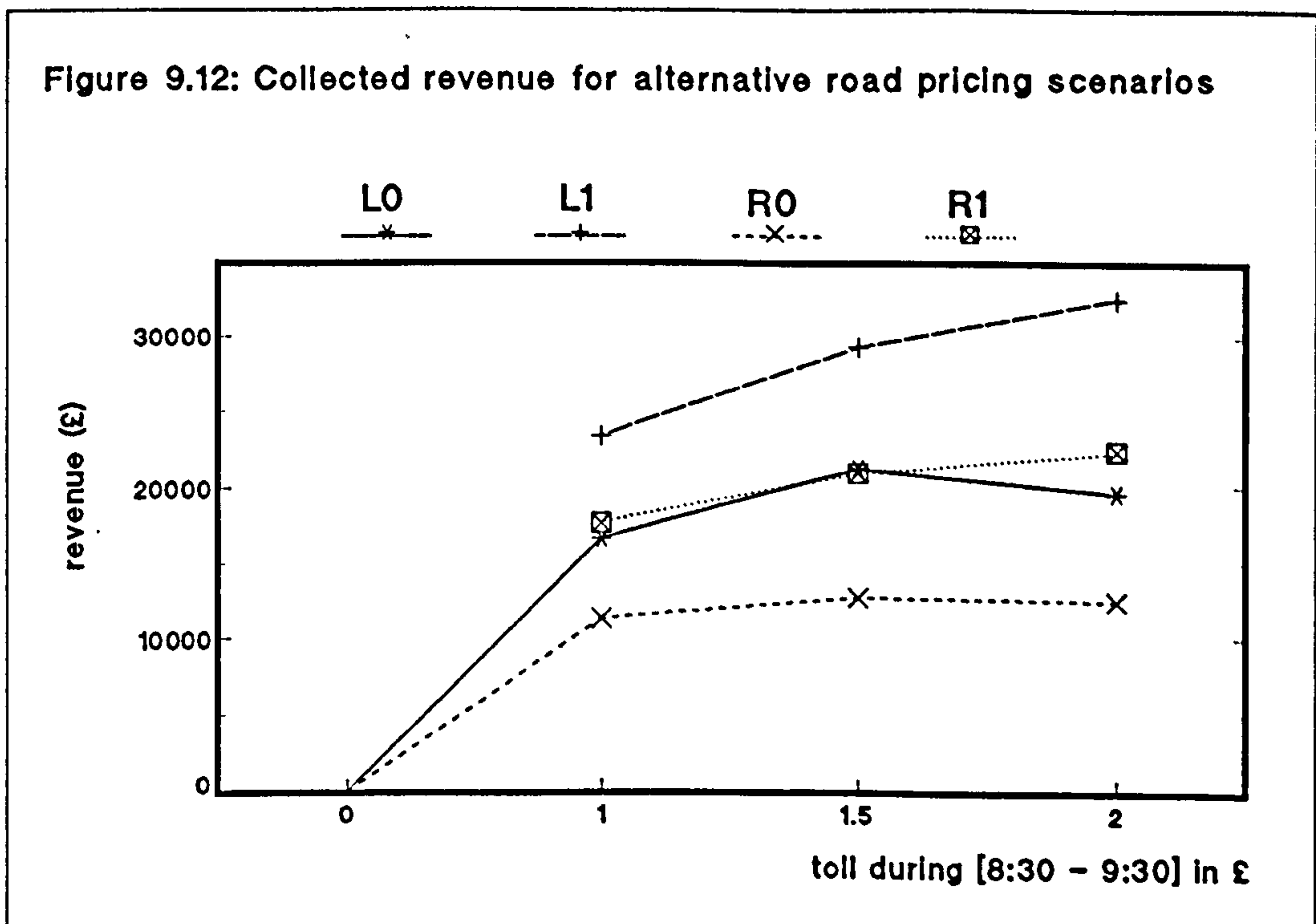






schedule delays. Early schedule delay is a monotonic increasing function of the toll charge since travellers are willing to arrive rather earlier than pay the toll. Figure 9.11 shows that late schedule delay is reduced as the level of charge increases. Such an outcome is supported by the fact that the toll period extends 15 mins after the latest on time arrival at work which is at 9:15 and therefore travellers who arrive late will also have to pay the toll, which is rather an unattractive option.

The revenue collected under the alternative pricing policies is illustrated in figure 9.12. Higher revenue is collected under the "links" schemes since a toll is charged in a larger number of links. When road pricing is implemented only during the period [8:30 9:30], it appears that a charge of £ 1.5 maximises the revenue, implying that the price elasticity is less than 1 up to the level of £ 1.5 beyond which further increases in the toll result in lower revenue. Figure 9.12 also shows that under the policies that implement a £ 1.0 charge during the time interval [8:00, 8:30] the revenue increases as the charge during the period [8:30, 9:30] increases. This, because drivers cannot shift to early departures that would imply no charge, i.e. before 8:00, since such departures would result in very high early schedule delays.



In the analysis presented in this section the travel demand was assumed to be inelastic, in other words it was assumed that travellers do not have the option of switching to alternative transport modes or not making the trip. A development of the DSUE model which incorporates these travel options is presented by Vythoukas (1990b).

### 9.3 Forecasting the effects of drivers information systems

The information provided to drivers who have access to driver information systems (DIS) may be descriptive or prescriptive. *Descriptive* systems provide information on the traffic conditions prevailing in the majority of links located within a section of the urban area that includes both the origin and the destination of the trip. *Prescriptive* systems make suggestions on the best available route to follow without providing any information on traffic conditions. Drivers who receive information about traffic conditions and recommended paths will not necessarily follow the suggestions. Instead, they will decide on which route to follow based on their own perceptions about traffic conditions, and the information they receive. The degree to which they will use the provided information depends on the reliability of the information system, the familiarity they have with the network, their experience on travelling at particular time intervals during the day and the specific circumstances that create the congestion. This, because traffic congestion in a network can be recurrent or caused by incidents. Recurrent congestion is due to the fact that the capacity of certain links is inadequate to serve the number of vehicles that want to use this particular link and follows similar patterns every day. On the other hand, incident congestion is caused by incidents that occur in the highway system, for example car breakdowns, accidents, etc, that temporarily reduce the capacity of a link. In the latter case it is more likely that the degree that DIS users will comply with the available information will be high since drivers cannot predict the duration of the incident and the way traffic congestion will evolve. In the former case, it is expected that drivers will base their route choice decisions on their experience, perceptions as well as the information they will receive.

This section uses the DSUE assignment model to forecast the effects of a descriptive information system and to estimate the expected benefits due to recurrent congestion only. The analysis draws heavily on the work by Koutsopoulos and Lotan (1990) who estimated the effects of driver information systems using static assignment procedures, hence without considering the time variability of traffic flow patterns. The analysis presented here represents the dynamics of traffic flow but also lies at a macro level and does not capture all the details of the problem. However it is useful for

understanding the interactions between the basic components of the problem and for facilitating a sensitivity analysis of the expected benefits.

The analysis of the impact of DIS on travel patterns is based on the hypothesis that drivers who receive information may alter their trip choices in the long run. This, since it is reasonable to assume that these drivers may realise that certain routes are consistently less congested than the ones that they usually follow, and that for certain departure times traffic conditions are consistently better. Thus, in developing the procedures that will be used to evaluate the impacts of DIS, certain assumptions are made with respect to drivers' trip choices and their response to information, and are presented below.

Following the general rule used in the development of dynamic assignment procedures, it is assumed that drivers select that combination of route and departure time which they believe, is associated with the minimum disutility. However, since they have imperfect knowledge on the actual traffic conditions in the network, they decide on which route to follow and what time to depart based on their perceptions on the travel times along the various routes connecting their origin to their destination, that they consider as reasonable options. Providing information on traffic conditions affects the perceptions that drivers have on the relative utility associated with alternative trip choices. Thus it is assumed that travellers who receive information will in the long run switch to routes and shift to departure times which actually perform better, while non informed drivers will continue using the routes and departure times which (based on their imperfect knowledge on traffic conditions) they believe will minimise their disutility of travel. Trip choices can then be modelled using the random utility framework, that was also used in the development of the DSUE assignment model.

Under the above assumptions the problem of estimating the effects of DIS on travel patterns can then be considered as a multi-class traffic assignment problem, in which two classes of users will be considered:

- The first consists of the drivers who have access to the information system. The provided information is assumed to be reliable, so that the majority of these travellers, after processing the information obtained from the DIS and the one developed by their own experience, will select one of the shortest routes recommended by the DIS. It is expected that all the drivers will not select the shortest route since it is reasonable to assume that for a number of travellers, particular attributes, which are different than travel time, may be the main determinants of their choices.

- the second class consists of drivers who do not have access to the DIS and make their trip choice decisions based solely on their own experience on traffic conditions and the information they might receive from media or other users of the network. Thus, these drivers are not expected to be concentrated only on the optimum alternative trip choices but to consider as alternative options routes which will not be used by the drivers who form the first class.

Within this conceptual framework, and in order to represent the influence of information on drivers' behaviour, it is assumed that drivers who belong to the first class are characterised by low variability of preferences with respect to route choice and therefore will be concentrated on the shortest travel time routes. On the other hand drivers who belong to the second category will be more evenly distributed over the optimum and non-optimum routes.

Thus, the two classes of users will be characterised by different parameters  $\mu_r$ . The class of travellers who are associated with low values of  $\mu_r$  is characterised by high variability of preferences with respect to route choice and can therefore represent the users who do not have access to the DIS system, or who do not use the provided information. Drivers characterised by high values of  $\mu_r$  represent the ones who can obtain information about the traffic conditions.

To estimate the potential benefits of route guidance and their sensitivity to various parameters, a number of experiments were carried out to simulate different scenarios of percentage of DIS users and level of congestion. The effect of the provided information on the form of the departure rate distribution will be analysed and average travel times will be estimated. The multiclass version of the DSUE model was applied to the Sioux Falls network; the demand adjustment mechanism in this case will not represent the evolution of the travel patterns from day to day, instead it will be used as an equilibration mechanism. In the simulation experiments that were carried out, the parameter that reflects the variability of preferences with respect to departure time has taken the same value for both classes of network users,  $\mu_{t1} = \mu_{t2} = 1.0$ ; regarding the parameter that reflects the variability of preferences with respect to route choice, for the DIS users it was considered to be  $\mu_{r1} = 10$  and for the drivers who do not have access to the DIS to be  $\mu_{r2} = 1.0$ .

Figure 9.13 illustrates the time dependent O-D travel times and departure rate distributions for the informed and non-informed drivers travelling from node 2 to node 19 for a 50% market penetration level of the DIS. The distributions presented

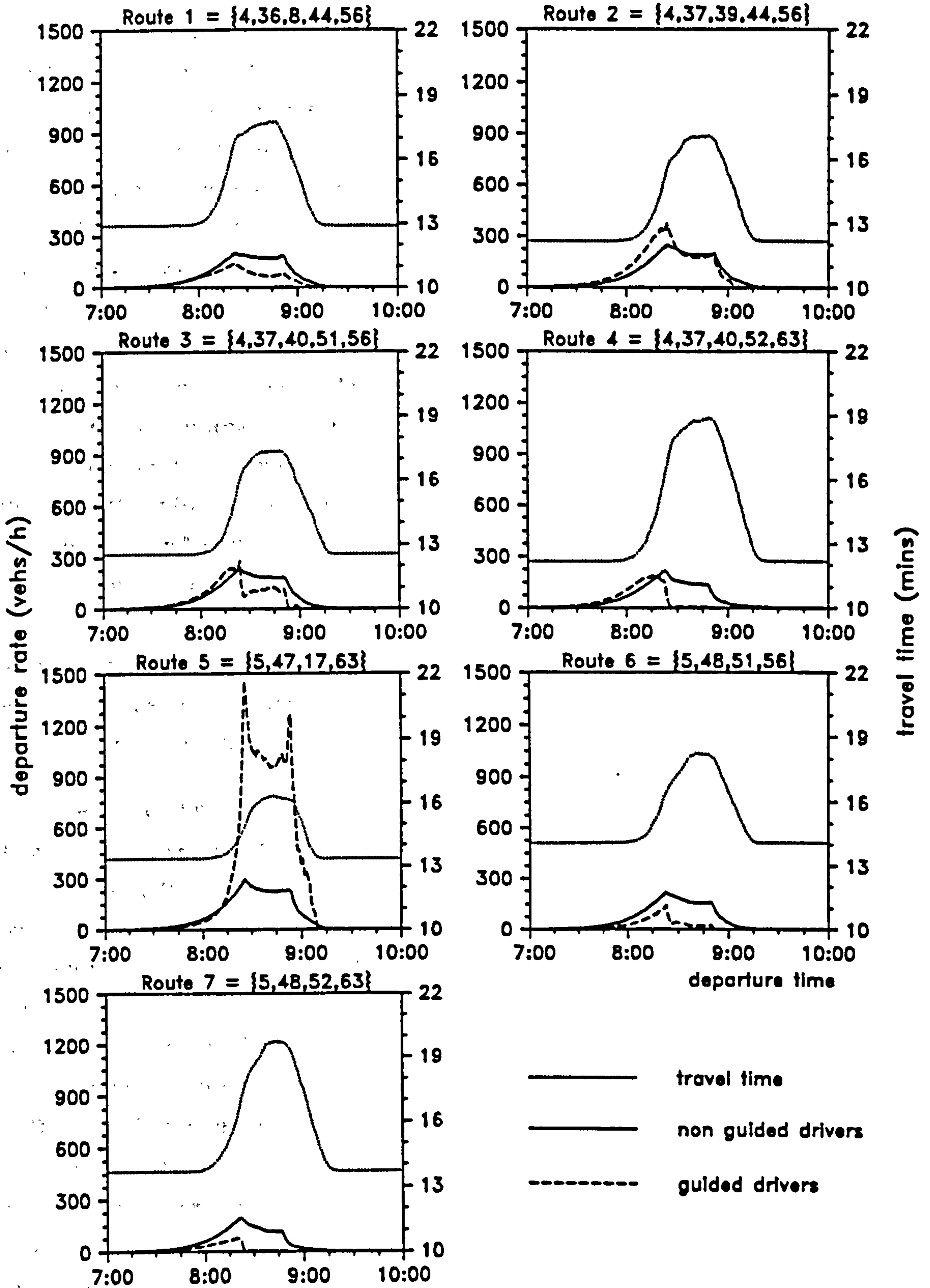


Figure 9.13: Travel time and departure rate distributions of informed and uninformed drivers for O-D pair 2-19 under 50% market penetration of the DIS.

correspond to the routes that are used by the guided vehicles; a very small proportion of the non-guided drivers follow other routes which are associated with longer travel times and are not presented in figure 9.13. The figure demonstrates that the highest proportion of the guided drivers follow route 5 which provide the better level of service, since it is the less congested (it has the lower peak). The highest peak on the departure rate distributions for both guided and non-guided drivers occur on route 5 just before 8:30 when the O-D travel time is about 15 mins which results in an almost on-time arrival at work. Guided vehicles also follow other non-optimum routes, but this is an outcome of the assumption that travellers will not necessarily follow exactly the optimum alternatives proposed by the DIS but will have better perceptions on travel conditions than the non guided drivers have. This is the reason why a much lower proportion of informed drivers follow the non-optimum routes. Furthermore, the analysis is based on the assumption that the network will reach a state of equilibrium, and therefore the information provided by the DIS may also direct drivers towards a set of "near optimum" routes rather than the shortest one, since if there was a higher concentration of drivers on route 5 this would create congestion, and route 5 would not be the optimum route. In general the figure illustrates that drivers who do not have access to the information are more uniformly distributed over the alternative reasonable routes, while guided drivers are more concentrated on those routes and departure times that are associated with shorter travel times.

The average travel times experienced by informed and non-informed drivers as a function of the percentage of drivers who receive information are depicted in figure 9.14. On average, travel times for the guided drivers increases as the percentage of equipped vehicles increases. Such an outcome is supported by the fact that, as information is available to more drivers, there will be a higher concentration of traffic (mainly consisting of equipped vehicles) on the optimum routes, which therefore become more congested. Travel times for the non-informed drivers are less affected and remain at an almost constant level with the minimum for a 50% market penetration of the DIS. The figure demonstrates that for the logit parameters  $\mu_{r1} = 10$  and  $\mu_{r2} = 1.0$  used in the simulation experiments, the travel times experienced by informed drivers are substantially lower than the ones experienced by non-informed drivers, and therefore the total travel time spent during the morning peak decreases as the number of the DIS users increases. Similar results were derived in Koutsopoulos's and Lotan's (1990) analysis, (in which static assignment procedures were used) with the difference that in their study average travel time for the non-guided vehicles was found to

Figure 9.14: Average travel times for guided and non guided drivers

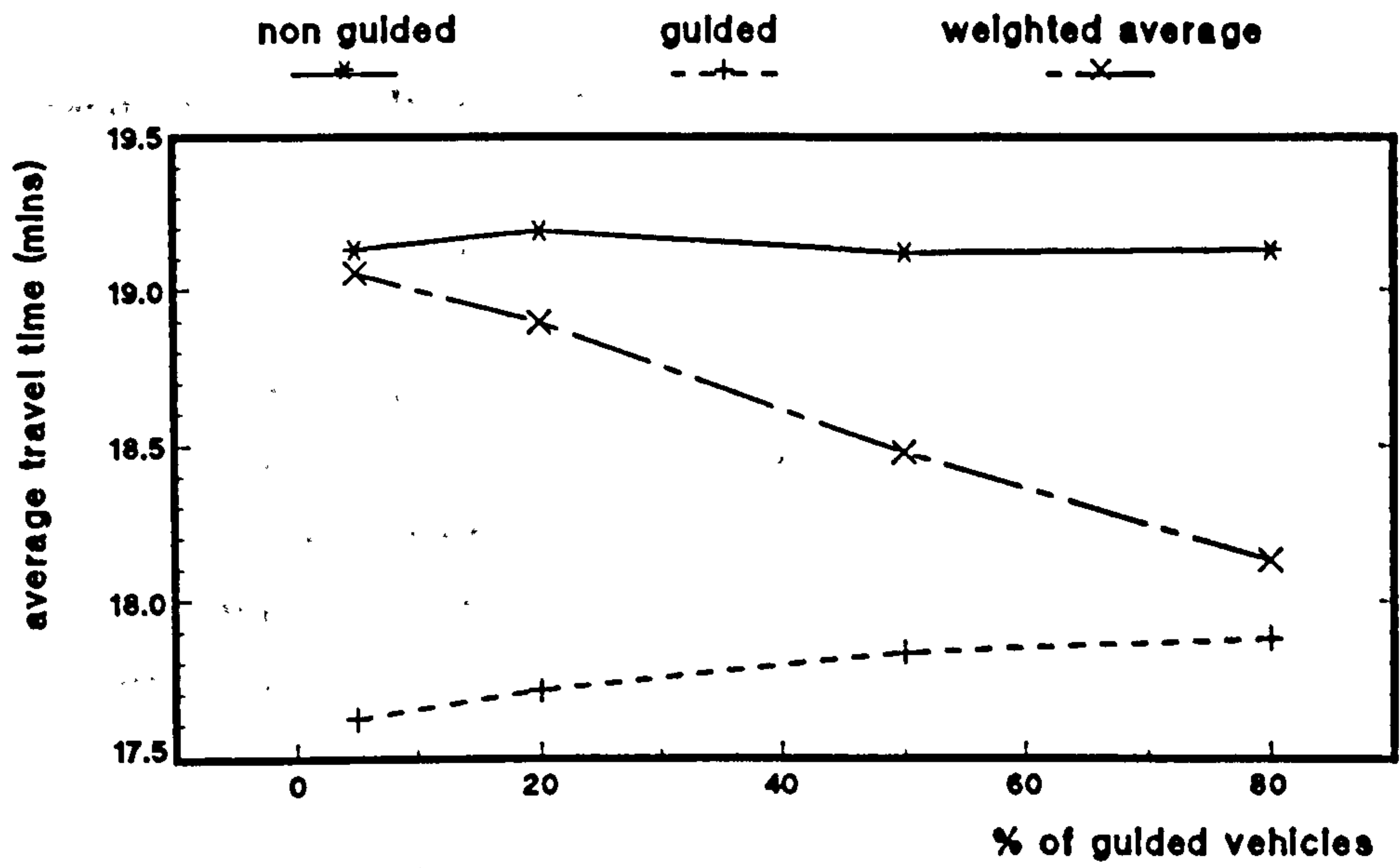
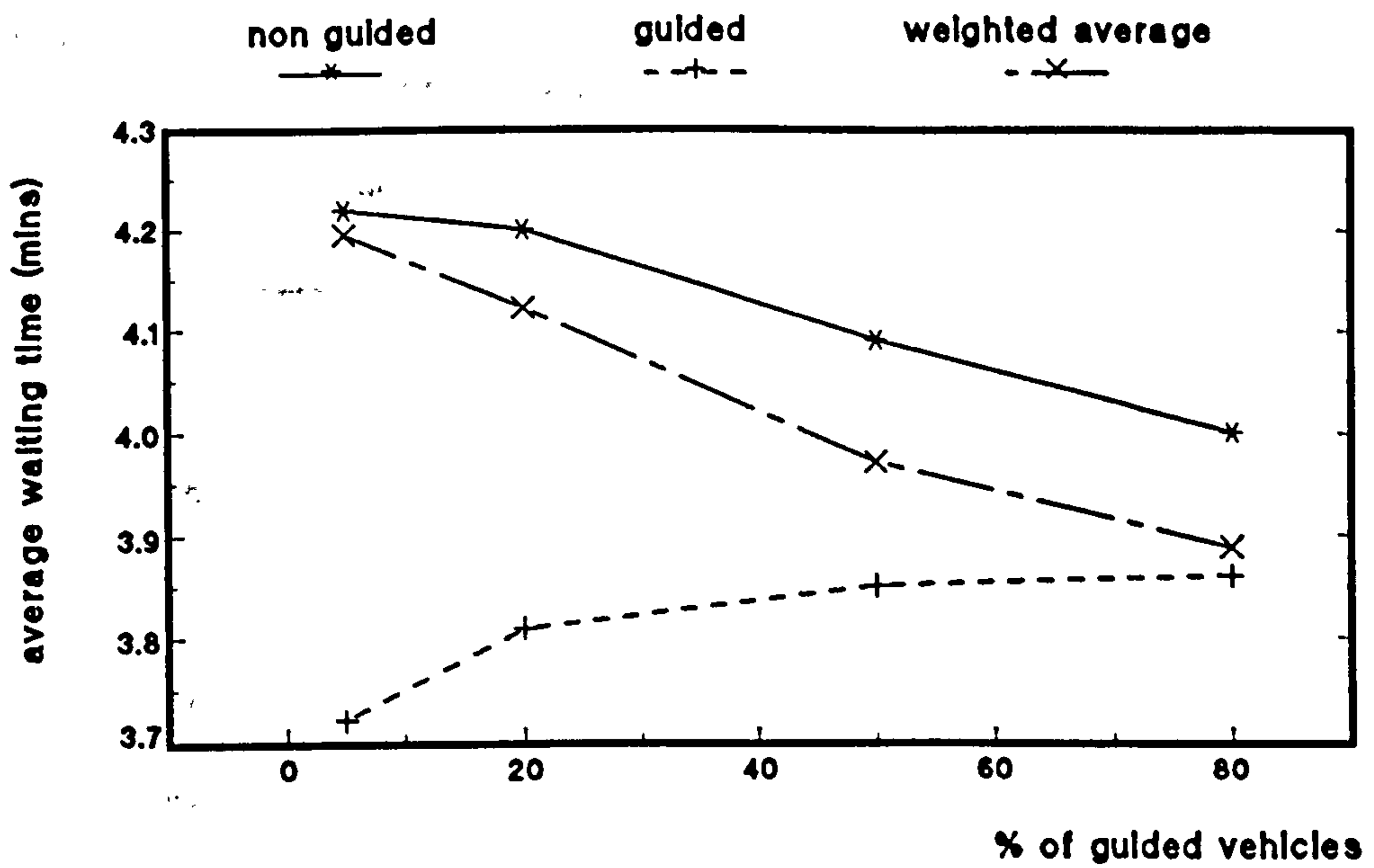


Figure 9.15: Average waiting times for guided and non guided drivers

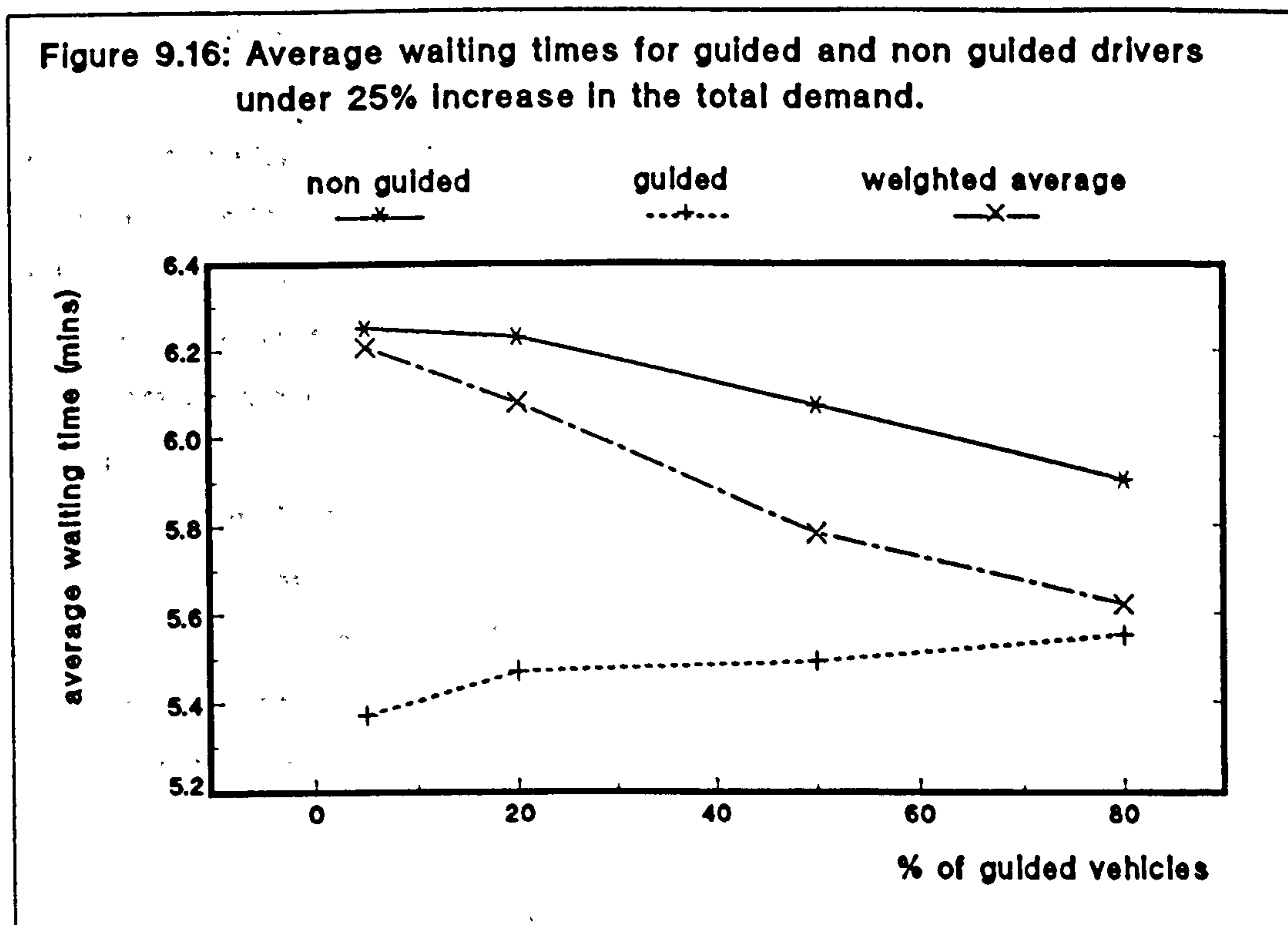




increase as the percentage of guided vehicles increases. However the level of increase found in their analysis is almost negligible (maximum increase is 0.02 mins).

Figure 9.15 illustrates the delays experienced by informed and uninformed drivers as a function of the percentage of informed drivers. This figure, in conjunction with figure 9.14 demonstrates that in the simulation experiments conducted, the benefits derived from the operation of the DIS are not so much attributable to savings in waiting time but rather to the diversion of drivers to shortest (when examined under free flow conditions) routes. This, because when the network capacity is limited and it has to accommodate demand which is concentrated (in terms of destinations and arrival times) in space and time, congestion occurs in the majority of the network links, and there are not enough links with excess capacity to accommodate that marginal demand which causes congestion. Thus, diverting traffic from one route to another will create congestion in the latter route; the proportion that will be diverted and the resulting level of congestion will depend on the overall level of congestion existing in the network.

For increased levels of congestion the behaviour of the system is generally the same. The saving in waiting times are higher as is shown in figure 9.16 that depicts the variation of waiting times for guided and non-guided vehicles under different market penetration levels of the DIS, and for 25% increase in the total level of demand.



Finally it should be noted that the difference in travel times between informed and uninformed drivers is a result of the values of the parameters  $\mu_{r1}$  and  $\mu_{r2}$  which have been used in the experiments but which have not been empirically tested.

#### 9.4 Summary

In this chapter the DSUE model has been used to estimate the effects that road pricing schemes and driver information systems will have on travellers trip choices and the resulting traffic patterns during the morning peak. As in chapter eight, the analysis is based on the results from a number of simulation experiments to evaluate the efficiency gains derived from various road pricing schemes and for different market penetration levels of the DIS.

The experiments have shown that drivers respond to road pricing by shifting to earlier departure times, in order to avoid the toll, or by switching to alternative routes which do not include links imposing a toll charge. As a result, i) the shifting peaks phenomenon may emerge, i.e. the development of a congestion period just before the toll period, and ii) a major switch of the demand from shorter "toll imposing" routes towards longer "no toll" routes may occur. For the "short duration" schemes, travel and waiting times are not monotonic functions of the toll charged. It appears that there is a certain level of toll that minimises the total travel time spent and the revenue collected during the morning peak. This since for lower levels of toll there is a certain proportion of travellers who are willing to pay the charge in order to avoid early schedule delays and experience better traffic conditions. However as the level of toll increases more travellers are willing to shift their trip before the toll period, although they will experience longer early schedule delays, instead of paying a very high toll. As a result high peaks may develop before the toll period, average travel time and early schedule delay increases and revenue decreases.

Similar phenomena are not observed when the toll period extends over a longer time period before the travellers' desired arrival times at their destinations. This since shifts to departures that imply no toll charge would result in very long early schedule delays. In general, under "longer duration" road pricing schemes, travel and waiting times decrease and revenue increases as the toll charges increase. Furthermore, for the same pricing scheme, (in terms of period of charge and level of toll) savings in travel and waiting times increase as the number of links imposing a toll charge increases.

To evaluate the effects that driver information systems will have on the traffic conditions during the morning peak, a multiclass version of the DSUE model was used. The two different classes considered are associated with different values of the parameter that represents the variability of preferences with respect to route choice, and are used in order to model the drivers who have access to the DIS and the ones who do not. Lower variability of preferences is associated with the informed drivers to represent the better perceptions on traffic conditions they have as a result of the information they acquired from the DIS.

Simulation experiments have shown that drivers who do not have access to the information system are more uniformly distributed over the alternative reasonable routes, while guided drivers are more concentrated on those routes that are associated with the shortest travel times. Furthermore, it was shown that on average the travel times experienced by the guided drivers increase as the percentage of the equipped vehicles increases. However the difference between the travel times experienced by guided and non guided drivers is substantial, so that the average travel time for all the network users decreases as the percentage of informed drivers increases. Examination of the drivers' waiting times have shown that travel time savings are more attributable to the diversion of guided drivers towards shortest (when examined under free flow conditions) routes rather than to savings in waiting times.

# 10 conclusions

The first part of the paper is devoted to the study of the asymptotic behavior of the solutions of the system (1) for large values of the parameter  $\epsilon$ . It is shown that the solutions of the system (1) are asymptotically equivalent to the solutions of the system (2) for large values of  $\epsilon$ . The asymptotic expansion of the solutions of the system (1) is obtained in the form of a power series in  $\epsilon^{-1}$ . The leading term of this expansion is the solution of the system (2). The higher order terms of the expansion are obtained by solving a sequence of linear systems. The asymptotic expansion of the solutions of the system (1) is valid for large values of  $\epsilon$  and for fixed values of  $t$  and  $x$ . The asymptotic expansion of the solutions of the system (1) is also valid for large values of  $\epsilon$  and for fixed values of  $t$  and  $x$  if the initial conditions are given at  $t=0$  and  $x=0$ . The asymptotic expansion of the solutions of the system (1) is also valid for large values of  $\epsilon$  and for fixed values of  $t$  and  $x$  if the initial conditions are given at  $t=0$  and  $x=0$  and the boundary conditions are given at  $x=0$  and  $x=L$ . The asymptotic expansion of the solutions of the system (1) is also valid for large values of  $\epsilon$  and for fixed values of  $t$  and  $x$  if the initial conditions are given at  $t=0$  and  $x=0$  and the boundary conditions are given at  $x=0$  and  $x=L$  and the system (1) is linear.

### 10.1 Summary and Conclusions

Defining the dynamic stochastic user equilibrium traffic patterns in an urban road network is a challenging problem and an important issue within the context of transportation planning analysis. Over the last decade considerable research effort has been focused on this topic, and there is still continuing interest and research on the development of procedures that will enable the dynamic traffic analysis of real networks. The increasing interest on dynamic network analysis has emerged in an attempt to improve the predictive capability of static assignment procedures which fail to represent the essential features of traffic congestion during peak periods since i) they do not represent the time variability of traffic patterns and ii) they do not model travellers' departure time choice. Both these decisions have an important role on the development of congestion patterns. On the other hand, existing departure time choice procedures model the time dimension of choice but do not represent the interactions between users and network performance and, existing dynamic stochastic assignment procedures model travellers' route and departure time choices and the resulting time varying traffic patterns, but restrict the topography of the network to specific simple forms and therefore cannot be used for the analysis of realistic networks.

The major achievement of this study was the development of a procedure which defines travellers' departure time and route choices and the resulting time varying traffic patterns during the morning peak, and which does not restrict the network topography to specific forms.

To facilitate the derivation of the DSUE assignment model a number of related research topics were also discussed. Thus, in order to comprehend the mechanisms directing travellers' route and departure time choice decisions the concepts and a review of the currently available basic models used in the analysis of multiple choice

behaviour was presented. The review was concentrated on the utility maximisation theory and particularly on the logit formulations. Furthermore, since the focus of this study lies within the broader context of traffic assignment, the necessary background information required for the analysis of the traffic assignment problems and, various static assignment modelling procedures were analysed. Furthermore the utility functions used in existing departure time choice models were presented and the currently developed dynamic network analysis models were also reviewed. These models take into account the spatial and time-of-day variability of network congestion but as was pointed out earlier they can only analyse specific network forms; the review provided the necessary background for developing the DSUE assignment model.

The developed model follows the framework of existing dynamic assignment models which analyse simple network forms, and thus consists of a travel time model, a demand model and a demand adjustment mechanism. The travel time model is used to calculate the time varying traffic patterns and travel times given the time dependent departure rates associated with the routes connecting the O-D pairs. The model has introduced the concept of reasonable paths under time varying traffic conditions, and the time dependent route choice set considered by the drivers. Furthermore, it models the complex interactions between the different flow components within any network form. Two alternative formulations of the travel time model were developed; the first is based on traffic flow theory and the second on queueing theory. The demand model is based on the utility maximisation decision rule and defines the time dependent departure rates following each reasonable route, given the total utility associated with each combination of departure time and route. The demand adjustment mechanism models the interaction between individuals' decisions as they are directed by their own criteria of choice and the transportation network performance; it describes the evolution of travel patterns from day to day and is used to derive the equilibrium solution. The mechanism has introduced the different classes of drivers who adjust their trip choices, in order to facilitate the convergence of the system towards the equilibrium state. Furthermore, a procedure based on successive averages of the time varying link flow patterns was proposed. The method provides an alternative framework for defining the equilibrium patterns. The advantage of this approach is that it does not require path enumeration, as the demand adjustment mechanism approach does and, it will therefore enable the analysis of large scale networks, if efficient dynamic stochastic network loading procedures are developed.

The model has demonstrated the ability to define the temporal distribution of traffic during the peak period; its ability to replicate this important characteristic of traffic

patterns is in contrast to static models that treat traffic flows as constant during the peak period congestion. Travel time distributions have a smooth shape, while departure rate distributions are continuous but do not have continuous derivatives. Numerical simulation experiments using the DSUE assignment model were conducted to analyse the impact that i) different work start time flexibilities, ii) different levels of variability of preferences, iii) different levels of demand and iv) increases in a bottleneck capacity, have on the peak period traffic patterns. The main conclusions derived from the experimental analysis are summarised below:

- The level of work start time flexibility has a major impact on the development of the congestion patterns during the morning peak. Flexible work start time policies allow travellers to arrive at their destination within a specific period of time during the morning peak and thus, the longer this period is, the wider the spread of the demand over time will be. The major effect of increasing the work start time flexibility is therefore the spreading of the departure rate distribution over a longer time period, which results in lower levels of congestion and lower O-D travel times. Under inflexible work start time policies, departure rate distributions exhibit a single high peak, reflecting the willingness of travellers to arrive at their destination at the same time; the highest peak is observed at that departure time that results in on-time arrivals at work and causes a further increase in the congestion level. In general, the study has confirmed the view that inflexible working times result in higher levels of congestion due to the high concentration of traffic within a short time interval while higher levels of work start time flexibilities result in lower travel times.
- The level of the variability of travellers' preferences which respect to route and departure time choice also affects the traffic patterns and the level of congestion developed in the transportation network. High variability of preferences, which reflects less accurate perceptions of the attributes the alternative combinations of routes and departure times available to the network users, results in a more uniform spread of the demand in space and time and consequently in lower levels of congestion and higher values of schedule delay (since there is a higher proportion of travellers who select non optimum routes and departure times). Travellers characterised by lower variability of preferences, have better perceptions of the alternatives' attributes and therefore their choices are more concentrated on the optimum routes and departure times. As a result higher levels of congestion are developed in the vicinity of the highest utility alternatives; travellers experience longer travel times, but in terms of the total utility of travel

they are better off since their trip choices result in much lower levels of schedule delay, than the ones experienced by travellers with high variability of preferences.

- Changes in the level of demand affect the level of congestion developed in the network. As expected, an increase in the overall level of demand results in higher levels of congestion. Peak period travel times may get substantially higher and the congestion period is extended over a longer period towards earlier departures. The shape of the travel time and departure rate distributions remain the same but as the demand increases the resulting distributions get more highly peaked. Furthermore, the mode of the departure rate distribution shifts to the left, revealing evidence that travellers respond to the higher level of congestion, resulting from the increased level of demand and the limited network capacity, by shifting towards earlier departure times.
- Increasing the capacity of a link affects the evolution of the congestion patterns during the morning peak, and may have different effects on the traffic conditions experienced by the drivers travelling between different O-D pairs. Simulation experiments have shown that a capacity expansion scheme may cause a shift of the demand towards later departures, since travellers who experience lower travel times (after the capacity has increased) would experience longer early schedule delays if they continue departing at the same time. Furthermore, the increase in a link capacity may reduce the level of congestion developed in that link, but it may also create congestion in bottlenecks located downstream. This is due to the fact that less traffic is held at this bottleneck and the increased level of flow that exits from this link and enters downstream links may be higher than the capacity of those links. Thus, queues may develop at previously non congested links, or increasing levels of congestion may appear in already congested links. The experiments have shown that, as a result of the development of new congestion patterns, drivers travelling between certain O-D pairs may on average experience better traffic conditions, while the average O-D travel time for other O-D pairs may be increased.

Furthermore, due to its inherent behavioural nature, the DSUE model has also the potential to evaluate a wide range of transportation policies not directly related to parameters it includes. To illustrate the potential of the model in this regard, various simulation experiments were carried out to assess the effects of various road pricing policies and driver information systems. The main conclusions from these experiments are summarised below:



- Travellers respond to road pricing by shifting towards earlier departures in order to enter the area where the pricing scheme is implemented before the start of the toll period. As a result the shifting peaks phenomenon may emerge, i.e., the development of a congestion period just before the toll period starts. Furthermore, a major switch from shorter "toll-imposing" routes to longer "no-toll" routes may occur, if the destination is located outside the toll ring. The experiments have shown that in general road pricing schemes reduce the level of congestion developed during the morning peak. However for "short duration" schemes, which extend only for a short time period before the desired arrival time at the destination, high congestion levels may develop before the toll period. This, because as the level of toll increases there is a greater proportion of travellers who are willing to arrive early instead of paying the toll and who create an earlier peak. Thus, for such policies there is an optimum toll which maximises the travel time and waiting time savings and also the revenue collected by the highway authorities. For "long-duration" schemes, a shift before the toll period would imply higher levels of schedule delay and thus there is a lower percentage of travellers who are not willing to pay the toll, at the cost of experiencing high early schedule delays. When a differentiated toll scheme is implemented two peaks are observed on traffic patterns, just before the toll period starts and just before the toll charge changes to a different level.
- Travellers who have access to a drivers information system are more concentrated on those routes which are associated with the shortest travel times, while uninformed drivers are more uniformly distributed over all the reasonable paths connecting their origin to their destination. On average the travel time experienced by the informed network users increases as the number of the equipped vehicles increases. However, the average travel time for all the network users decreases as the percentage of informed users increases since, (for the values of the parameters used in the simulation experiments) the difference between travel times experienced by informed and uninformed users is high. Furthermore, in the experiments conducted, the lower travel times experienced by the informed drivers are not so much attributed to savings in waiting times but rather to the diversion of these drivers to short routes.

To summarise, the major achievement of this study is the development of a model with the ability to;

- define travellers' route and departure time choices and the resulting temporal distribution of traffic during the morning peak, in contrast to static assignment models which consider traffic flows as constant during the peak,
- analyse any network form in contrast to existing dynamic stochastic assignment models which can handle only specific network forms,
- replicate the phenomenon of spreading peaks in modelling various policies such flexible work starting time strategies, road pricing, etc.
- estimate the potential increase of congestion levels in links downstream of an expanded capacity bottleneck, which cannot be estimated by static assignment procedures,
- evaluate the impact of various road pricing schemes on the evolution of traffic congestion patterns,
- assess the effects of drivers information systems on the travel time experienced by guided and unguided drivers,

## 10.2 Limitations and further research.

As an initial effort towards the development of procedures that solve the problem of estimating route and departure time choices and the resulting traffic patterns during the morning peak and, given that research in this area is still at its initial stage, a number of aspects of the problem were not examined in depth and more research both at the theoretical and empirical level is still needed. Below the limitations of this study along with recommendations for further research are discussed.

- In the proposed traffic flow model traffic patterns are represented using continuous path flow profiles as opposed to packets of vehicles that are also used in dynamic traffic assignment procedures. Clearly, a packet based approach can provide a better representation of the traffic patterns especially when the size of the packet is small; at the extreme case, simulating individual vehicles would be the ideal way of representing traffic. However such approaches require more elaborate procedures in order to keep track of the packets in space and time, and thus impose high computational costs. To overcome this problem the size of the packet must be increased, which however eliminates the advantages of the packet based approach.
- During a time slice the traffic conditions within a link are assumed uniform. This might not be a realistic assumption, but when the time slice is short (as in the

applications presented in this study) it can be a good approximation of the real world traffic conditions. Furthermore, capturing the space and time dynamics of flow within a link and during each time slice would result in prohibitive computational complexity and costs.

- In developing two major components of the DSUE assignment model, namely the demand model and the demand adjustment mechanism, a logit formulation is adopted. As a result, the undesirable property of the independence of the irrelevant alternatives (IIA) is inherent in the DSUE model. However, since the attributes of the alternative routes (e.g. travel time) are dependent on the level of demand, the effects of the IIA property are neutralised, as was also pointed out by Fisk (1980) in static assignment formulations. An alternative approach to the problem of modelling travellers' route and departure time choices would be to use probit formulations which however are associated with high computational costs. Although research in the area of estimating the choice probabilities derived from probit models receives considerable attention (McFadden 1990), it is rather computationally unfeasible to use current procedures in traffic assignment that solve the choice probabilities using probit models.
- An area that requires further research deals with the formulation of the link performance functions and their use in the dynamic assignment procedures. Clearly a queueing theory based travel time model provides a more realistic representation, since it has the ability to replicate the evolution of queues and the delays associated with them. However in such a procedure, the travel time needed to traverse a link should not be constant, but flow dependent. Alternatively, if a travel time model based on traffic flow theory is applied, the link performance functions must model delays as derived from a deterministic time dependent queueing model and must also take into account the effects of traffic density on link travel times.
- Further research is required to develop more realistic procedures which will model the day to day adjustment of travellers' choice. Empirical work is required in order to identify the nature of information acquired by the travellers, and the mechanisms used by them in order to process this information in conjunction with their own experiences.
- The proposed model is dynamic since it models the within the day variability of travel demand, but it is also equilibrium oriented. However it is still uncertain whether equilibria do exist in real world urban road networks. Of course the day to day variability of traffic patterns can be attributed to the different levels of demand that use the network each day, and an equilibrium approach can be used

as a means to define an average state of the system. Nevertheless, empirical work is also required to test the capacity of existing dynamic assignment procedures to represent traffic flow patterns and therefore identify the limitations in the assumptions used in the development of such procedures.

- At its current form, the model requires high computational resources as compared to static assignment procedures. This, however, should not be considered as a major limitation of the model, since it calculates a wider range of variables, such as time varying traffic flow profiles and time dependent departure rate distributions for each O-D pair. Furthermore, since the purpose of this study was to develop a research tool rather than a commercial software package, further work on the development of the software might considerably improve the efficiency of the model and enable the application of the model to large scale networks.

# references

- Abdulaal M. and LeBlank L.J. (1979). "Continuous Equilibrium Network Design Models". *Transportation Research 13B*, pp. 19-32
- Abkowitz M.D. (1980). *The Impact of Service Reliability on Work Travel Behavior*, Ph.D. Dissertation, Department of Civil Engineering, Massachusetts Institute of Technology.
- Abkowitz M.D. (1981a) "An Analysis of the Commuter Departure Time Decision". *Transportation Vol. 10*, PP. 283-297.
- Abkowitz M.D. (1981b) "Understanding the Effect of Transit Service Reliability on Work Travel Behavior". *Transportation Research Record 794*, pp. 33-41.
- Akcelik R. (1978). "A New Look at Davidson's Travel Time Function", *Traffic Engineering and Control, Vol. 19*, pp. 459-463.
- Alfa A.S. (1981). "Time-Dependent Route Assignment of Peak Traffic". *Transportation Engineering Journal, ASCE, Vol. 107*, pp. 153-165.
- Alfa A.S. (1982). "Individual Strategies for Combating Congestion". *Naval Research Logistics Quarterly, Vol. 29, No. 3*, pp. 475-482.
- Alfa A.S. (1986). "A Review of Models for the Temporal Distribution of Peak Traffic Demand". *Transportation Research Vol 20B, No. 6*, pp. 491-499.
- Alfa A.S. and Minh D.L. (1979). "A Stochastic Model for the Temporal Distribution of Traffic Demand - The Peak Hour Problem", *Transportation Science Vol. 13, No. 4*, pp. 315-324.
- Beckmann M.J., McGuire C.B. and Winsten C.B. (1956). *Studies in the Economics of Transportation*. Yale University Press, New Haven.
- Ben-Akiva M. (1973). "Structure of Passenger Travel Demand Models", Ph.D Thesis, Department of Civil Engineering, M.I.T.
- Ben-Akiva M. and Lerman S.R. (1979). "Disaggregate Travel and Mobility-Choice Models and Measures of Accessibility". In *Behavioural Travel Modelling*. D.A. Hensher and P.R. Stopher, eds. Croom Helm, London.
- Ben-Akiva M. (1985). "Dynamic Network Equilibrium Research". *Transportation Research Vol 19A, No. 5/6*, pp. 429-431.
- Ben-Akiva M. and Watanada T. (1981). "Application of a Continuous Spatial Choice Logit Model". In *Structural Analysis of Discrete Data with Econometric Applications*, edited by C.F. Manski and D. McFadden, M.I.T. Press, Cambridge, Mass.
- Ben-Akiva M., Cyna M. and de Palma A. (1984). "Dynamic Model of Peak Period Congestion". *Transportation Research, Vol. 18B, No. 4/5*, pp. 339-355.
- Ben-Akiva M. and Lerman S.R. (1985). *Discrete Choice Analysis: Theory and Application to Travel Demand*. The MIT Press, Cambridge, Massachusetts.
- Ben-Akiva M., Litinas N. and Tsukonawa K. (1985). "Continuous Spatial Choice : The Continuous Logit Model and Distributions of Trips and Urban Densities". *Transportation Research, Vol. 19A, No. 2*.

- Ben-Akiva M. and de Palma A. (1986). "Analysis of a Dynamic Residential Location Choice Model with Transaction Costs". *Journal of Regional Science Vol. 26, No. 2*.
- Ben-Akiva M. and de Palma A. (1987). "Dynamic Models of Transportation Networks", *Proceedings of Seminar D - Transportation Planning Methods, pp. 233-245, PTRC Summer Annual Meeting, University of Bath*.
- Ben-Akiva M., de Palma A., and Kanaroglou P. (1986a). "Dynamic Model of Peak Period Traffic Congestion with Elastic Arrival Rates". *Transportation Science Vol 20, No. 2, pp. 164-181*.
- Ben-Akiva M., de Palma A., and Kanaroglou P. (1986b). "Effects of Capacity Constraints on Peak-Period Traffic Congestion". *Transportation Research Record 1085, pp. 16-26*.
- Blum J.R. (1954). "Multidimensional Stochastic Approximation Methods". *Annals of Mathematical Statistics 25, pp737-744*.
- Bovy P.H.L and Jansen G.R.M. (1983). "Network Aggregation Effects upon Equilibrium Assignment Outcomes : An Empirical Investigation". *Transportation Science Vol. 17, No. 3, pp. 240-262*.
- Boyce D.E. (1984). "Urban Transportation Network-Equilibrium and Design Models : Recent Achievements and Future Prospects". *Environment and Planning A, Vol 16, pp. 1445-1474*.
- Boyce D.E., Janson B.N. and Eash R.W. (1981). "The Effect on Equilibrium Trip Assignment of Different Link Congestion Functions". *Transportation Research Vol. 15A, pp. 223-232*.
- Branston D. (1976). "Link Capacity Functions: A Review". *Transportation Research Vol. 10, pp. 223-236*.
- Bureau of Public Roads (1950). *Highway Capacity Manual, Practical Applications of Research*. Committee on Highway Capacity, Dep. of Traffic and Operations, Highway Research Board, U.S Dep. of Commerce.
- Bureau of Public Roads (1964). *Traffic Assignment Manual*. U.S. Department of Commerce, Washington, D.C.
- Burrell J.E. (1968). "Multipath Route Assignment and its Application to Capacity Restraint." *Proceedings of the 4th International Symposium on the Theory of Road Traffic Flow, Karlsruhe, West Germany*.
- Burrell J.E. (1976). "Multipath Route Assignment: A Comparison of Two Methods". In M. Florian (ed.), *Traffic Equilibrium Methods, Lecture Notes in Economics and Mathematical Systems 118, Springer-Verlag, New York, pp. 229-239*.
- Carey M. (1986). "A Constraint Qualification for a Dynamic Traffic Assignment Model". *Transportation Science Vol. 20, No. 1, pp. 55-58*.
- Carey M. (1987). "Optimal Time-Varying Flows on Congested Networks". *Operations Research, Vol. 35, No. 1, pp. 58-69*.

- Carey M. and Srinivasan A. (1987). "Congested Network Flows: Time-Varying Demands and Start-Time Policies". Forthcoming in the *European Journal of Operational Research*.
- Catling I. (1977). "Development of Junction Delay Formula" *LTRI Working paper 2*.
- Catling I. and Harbord B.J. (1985). "Electronic Road Pricing in Hong Kong: 2. The Technology". *Traffic Engineering and Control Vol. 26, No. 12, pp. 608-615*.
- Chang S.S.L. (1977). "Simulation of Transient and Time Varying Conditions in Queueing Networks". *Proceedings of the 8th Annual Pittsburgh Conference on Modelling and Simulation, Vol. 8, Part 11, pp. 1075-1078, Instrument Society of America*.
- Chang G.-L., Mahmassani H.S. and Herman R. (1985). "Macroparticle Traffic Simulation Model to Investigate Peak-Period Commuter Decision Dynamics". *Transportation Research Record 1005, pp. 107-121*.
- Chang G.-L. and Mahmassani H.S. (1988). "Travel Time Prediction and Departure Time Adjustment Behaviour Dynamics in a Congested Traffic System". *Transportation Research Vol. 22B, No. 3, pp. 217-232*.
- Clark C.E. (1961). "The Greatest of a Finite Set of Random Variables". *Operations Research Vol. 9, No. 2*.
- Cosslett S. (1977). The Trip Timing Decision for Travel to Work by Automobile. Chapter 3 (Part II) in *The Urban Travel Demand Forecasting Project Volume V: Demand Model Estimation and Validation*, by D. McFadden, A. Talvitie, S. Cosslett, I. Hasan, M. Johnson, F. Reid and K. Train, University of California-Berkeley.
- Dafermos S. (1971). "An Extended Traffic Assignment Model with Applications to Two-Way Traffic". *Transportation Science Vol. 5, pp. 366-389*.
- Dafermos S. (1980). "Traffic Equilibrium and Variational Inequalities". *Transportation Science Vol. 14, No. 1, pp. 42-54*.
- Dafermos S. (1982). "Relaxation Algorithms for the General Assymmetric Traffic Equilibrium Problem". *Transportation Science Vol. 16, No. 2, pp. 231-240*.
- Dafermos S.C. and Sparrow F.T. (1969). "The Traffic Assignment Problem for a General Network". *Journal of Research of the National Bureau of Standards - Mathematical Sciences*.
- Daganzo C.F. (1977a). "On the Traffic Assignment Problem with Flow Dependent Costs - I". *Transportation Research Vol. 11, pp. 433-437*
- Daganzo C.F. (1977b). "On the Traffic Assignment Problem with Flow Dependent Costs - II". *Transportation Research Vol. 11, pp. 439-441*
- Daganzo C.F. (1979). *Multimodal Probit : The Theory and Its Application to Demand Forecasting*. Academic Press, New York.
- Daganzo C.F. (1980a). "An Equilibrium Algorithm for the Spatial Aggregation Problem of Traffic Assignment". *Transportation Research Vol 14B, pp. 221-228*.



- Daganzo C.F. (1980b). "Network Representation, Continuum Approximation and a Solution to the Spatial Aggregation Problem of Traffic Assignment". *Transportation Research Vol 14B*, pp. 229-239.
- Daganzo C.F. (1983). "Stochastic Network Equilibrium with Multiple Vehicle Types and Assymmetric, Indefinite Link Cost Jacobians". *Transportation Science Vol. 17, No. 3*, pp. 282-299.
- Daganzo C.F. (1985). "The Uniqueness of a Time-dependent Equilibrium Distribution of Arrivals at a Single Bottleneck". *Transportation Science Vol. 19, No. 1*, pp. 29-37.
- Daganzo C. F. and Sheffi Y. (1977). "On Stochastic Models of Traffic Assignment". *Transportation Science Vol. 11, No. 3*.
- Daganzo C.F., Bouthelie F. and Sheffi Y. (1977). "Multinomial Probit and Qualitative Choice : A Computational Efficient Algorithm". *Transportation Science Vol. 11, No. 4*.
- Daly A. and Zachary S. (1978). "Improved multiple choice models", in *Determinants of Travel Choice*, Hensher D.A. and Dalvi M.Q. (eds.), Saxon House, Farnborough, England.
- Davidson K.B. (1966). "A Flow Travel-Time Relationship for Use in Transportation Planning". *Proceedings of the 3rd Australian Road Research Board Conference, Vol. 3*, pp. 183-194.
- Davidson K.B. (1978). "The Theoretical Basis of a Flow Travel-Time Relationship for Use in Transportation Planning". *Australian Road Research Vol. 8*, pp. 32-35.
- Dawson J.A.L. and Brown F.N. (1985). "Electronic Road Pricing in Hong Kong: 1 A Fair Way to Go ? ". *Traffic Engineering and Control Vol. 26, No. 11*, pp. 522-525
- Deneubourg J.L., de Palma A. and Kahn D. (1979). "Dynamic Models of Competition between Transportation Modes". *Environment and Planning, Vol. 11A*, pp. 665-673.
- de Palma A. and Lefevre C. (1983). "Individual Decision-Making in Dynamic Collective Systems ". *Journal of Mathematical Sociology Vol. 9*, pp. 103-124.
- de Palma A., Ben-Akiva M., Lefevre C. and Litinas N. (1983). "Stochastic Equilibrium Model of Peak Period Traffic Congestion". *Transportation Science Vol. 17, No. 4*, pp. 430-453.
- de Palma A., Ingenbleek J.-F., Lefevre C. and Ben-Akiva M. (1984). "Dynamic Model of Traffic Congestion in a Corridor". *Unpublished paper, Universite Libre de Bruxelles*.
- D'Este G. (1985). "The Effect of Staggered Working Hours on Commuter Trip Durations". *Transportation Research Vol 19A, No. 2*, pp. 109-117.
- Dial R.B. (1971). "A Probabilistic Multipath Traffic Assignment Algorithm which Obviates Path Enumeration". *Transportation Research Vol. 5, No. 2*, pp. 83-111.
- Domencich T. and McFadden D. (1975). *Urban Travel Demand - A Behavioural Analysis*. North Holland, Amsterdam.
- Dow P.D.C. (1979). "Capacity Restrained Road Assignment, III". *Traffic Engineering and Control Vol. 20*, pp. 303-305.

Eash R.W., Janson B.N. and Boyce D.E. (1979). "Equilibrium Trip Assignment: Advantages and Implication for Practice". *Transportation Research Record* 728, pp. 1-7.

Eash R.W., Chon K.S., Lee Y.J. and Boyce D.E. (1983). "Equilibrium Traffic Assignment as an Aggregate Network for Sketch Planning". *Transportation Research Record* 944, pp. 30-37.

Evans S.P. (1976). "Derivation and Analysis of some Models for Combining Trip Distribution and Assignment". *Transportation Research Vol. 10* pp. 37-57.

Fan W. (1976). "Simulation of Queueing Network with Time Varying Arrival Rates". *Transactions of the International Association for Mathematics and Computers in Simulation, Vol. XVIII, No. 3*, pp. 165-170.

Fargier (1981). "Effects of the Choice of Departure Time on Road Traffic Congestion". *Proceedings of the Eighth International Symposium on Transportation and Traffic Theory, Toronto, Canada*.

Fisk (1980). "Some Developments in Equilibrium Traffic Assignment". *Transportation Research Vol. 14B., No. 3*, pp. 243-255.

Fisk C. and Nguyen S. (1982). "Solution algorithms for Network Equilibrium Models with Assymmetric User Costs". *Transportation Science Vol. 16, No. 3*, pp. 361-381.

Florian M. (1984). "An Introduction to Ntework Models used in Transportation Planning". *Transportation Planning Models, Proceedings of the Course given at Ammalfi, Italy, Oct. 1982, (ed. M.Florian) North-Holland*, pp. 137-152.

Florian M. and Fox B. (1976). "On the Probabilistic Origin of Dial's Multipath Traffic Assignment Model". *Transportation Research Vo. 10, No. 5*, pp. 339-341.

Florian M. and Nguyen S. (1976). "Recent Experince with Equilibrium Methods for the Study of a Congested Area". *Lecture Notes in Economics and Mathematical Systems, 118 : Traffic Equilibrium Methods, Montreal 1974., Ed. M. Florian, Springer-Verlag, Berlin*, pp. 382-395.

Ford and Fulkerson (1962). "Flows in Networks" *Princeton University Press, Princeton, New Jersey*.

Frank M. and Wolfe P. (1956). "An Algorithm for Quadratic Programming". *Naval Research Logistics Quarterly Vol. 3*, pp. 95-110.

Friesz T.L. (1985). "Transportation network Equilibrium, Design and Aggregation: Key Developments and Research Opportunities". *Transportation Research Vol. 19A, No. 5/6*, pp. 413-427.

Fukushima M. (1984). "A Modified Frank-Wolfe Algorithm for Solving the Traffic Assignment Problem". *Transportation Research Vol. 18B, No.2*, pp. 169-177.

Gartner N.H., Gershwin S.B., Little J.D.C. and Ross P. (1980). "Pilot Study of Computer-Based Traffic Management ". *Transportation Research, Vol. 14B*, pp. 203-217.

- Haghani A.E. and Daskin M.S. (1983). "A Network Design Application of an Extraction Algorithm for Network Aggregation". *Transportation Research Record* 944, pp. 37-46.
- Hammerslag R. (1988). "A Threedimensional Assignment in the Time-Space". *Twentieth Annual UTSG Conference, University College London, Jan. 1988*.
- Hartgen D.T. (1974). "A Dynamic Model of Travel Mode Switching Behaviour". *Transportation Vol. 3, No. 1*.
- Hartley T.M. and Ortuzar J. de D. (1980). "Aggregate Modal Split Models : Is Current U.K. Practice Warranted ? ". *Traffic Engineering and Control Vol. 21, No. 1*.
- Hearn D.W. (1976). "Practical and Theoretical Aspects of Aggregation Problems in Transportation Planning Models". *Lecture Notes in Economics and Mathematical Systems, 118 : Traffic Equilibrium Methods, Montreal 1974., Ed. M. Florian, Springer-Verlag, Berlin, pp. 257-287*.
- Hendrickson C. and Kocur G. (1981). "Schedule Delay and Departure Time Decisions in a Deterministic Model". *Transportation Science Vol. 15, No. 1, pp. 62-77*.
- Hendrickson C., Nagin D. and Plank E. (1981). "Characteristics of Travel Time and Dynamic User Equilibrium for Travel to Work". *Proceedings of the Eighth International Symposium on Transportation and Traffic Theory, Toronto, Canada*.
- Hendrickson C. and Plank E. (1984). "The Flexibility of Departure Times for Work Trips". *Transportation Research Vol 16A, No. 1, pp. 25-36*.
- Herman R. and Lam T. (1974). "Trip Time Characteristics of Journeys to and From Work". *Proceedings of the Sixth International Symposium on Transportation and Traffic Flow, Sydney, Australia*.
- Heydecker B.G. (1983). "Some Consequences of Detailed Junction Modeling in Road Traffic Assignment". *Transportation Science Vol. 17, No. 3, pp. 263-281*.
- Ho J.K. (1980). "A Successive Linear Optimisation Approach to the Dynamic Traffic Assignment Problem". *Transportation Science Vol. 14, No. 4, pp. 295-305*.
- Horowitz J.L. (1984). "The Stability of Stochastic Equilibrium in a Two-Link Transportation Network". *Transportation Research Vol. 18B, No. 1, pp. 13-28*.
- Horowitz J.L., Sparmann J.M. and Daganzo C.F. (1982). "An Investigation of the Accuracy of the Clark Approximation for the Multinomial Probit Model". *Transportation Science Vol. 16, No. 3*.
- Hurdle (1974). "The Effect of Queueing on Traffic Assignment in a Simple Road Network". *Proceedings of the Sixth International Symposium on Transportation and Traffic Theory, Sydney, Australia, pp. 519-540*.
- Hurdle V. (1986). "On a Paper by Andre de Palma, Moshe Ben-Akiva, Claude Lefevre, and Nicolaos Litinas Entitled 'Stochastic Equilibrium Model of Peak Period Traffic Congestion' ". *Transportation Science Vol. 20, No. 4, pp. 287-289*.

Jones P. (1989). "The Restraint of Road Traffic in Urban Areas, Objectives, Options and Experiences", Rees Jeffreys' discussion paper No. 3, Transport Studies Unit, University of Oxford.

Jovanis P.P. and Moore A. (1984). "Models of Employee Work Schedule". *Proceedings of the Ninth International Symposium on Transportation and Traffic Theory, Delft University, Holland.*

Kemp M.A. (1977). "Management of Demand". *Transportation System Management, Transportation Research Board Special Report 172, National Academy of Sciences, Washington, D.C*

Koutsopoulos H.N. and Lotan T. (1990). "Effectiveness of Motorist Information Systems in Reducing Traffic Congestion". Conference Record of Papers presented at the First Vehicle Navigation and Information Conference, Toronto, Canada, 1989

Kraft G. and Wohl M. (1967). "New Directions for Passenger Demand Analysis and Forecasting". *Transportation Research Vol 1, pp. 205-230.*

Kroes E.P., Antonisse R.W. and Bexelius S. (1987). "Return to the Peak ?". *Proceedings of Seminar D - Transportation Planning Methods, pp. 233-245, PTRC Summer Annual Meeting, University of Bath.*

Kurtz T.G. (1978). "Strong Approximation Theorems for Density Dependent Markov Chains ". *Stochastic Processes and Their Applications, Vol. 6, pp. 223-240.*

Kuwahara M. and Newell G.F. (1987). "Queue Evolution on Freeways Leading to a Single Core City During the Morning Peak". *Proceedings of the Tenth International Symposium on Transportation and Traffic Flow Theory, M.I.T., pp. 21-40.*

Lam W.H.K. (1988). "Effects of Road Pricing on System Performance". *Traffic Engineering and Control Vol. 19, No. 12, pp. 631-635.*

Langdon M.G. (1984a). "Methods of Determining Choice Probability in Utility Maximising Multiple Alternative Models". *Transportation Research Vol. 18B, No. 3.*

Langdon M.G. (1984b). "Improved Algorithms for Estimating Choice Probabilities in the Multinomial Probit Model". *Transportation Science, Vol. 18, No. 3.*

LeBlank L.J. (1975). "An Algorithm for the Discrete Network Design Problem". *Transportation Science Vol. 9, pp.183-199*

LeBlank L.J., Morlok E.K. and Pierskalla W.P. (1975). "An Efficient Approach to solving the Road Network Equilibrium Traffic Assignment Problem". *Transportation Research Vol. 9, pp. 309-318.*

Leonard D.R., Tough J.B. and Baguley P.C. (1978). "CONTRAM : A Traffic Assignment Model for Predicting Flows and Queues during Peak Periods". *TRRL Laboratory Report 841.*

Lerman S.R and Manski C.F (1978). "An Estimator for the Generalised Multinomial Probit Choice Model". *56th Annual TRB Meeting, Washington, D.C.*

Lisco T.E. (1983). "Procedure for Predicting Queues and Delays on Expressways in Urban Core Areas". *Transportation Research Record 944, pp. 148-154.*

Lupi M. (1986). "Convergence of the Frank-Wolfe Algorithm in Transportation Network". *Civil Engineering Systems, Vol. 3, pp. 7-15.*

Magnanti T.L. (1984). "Models and Algorithms for Predicting Urban Traffic Equilibria". *Transportation Planning Models, Proceedings of the Course given at Amalfi, Italy, Oct. 1982, (ed. M.Florian) North-Holland, pp. 137-152.*

Mahmassani H.S. and Herman R. (1984). "Dynamic User Equilibrium Departure Time and Route Selection on Idealised Traffic Arterials". *Transportation Science Vol. 18, No. 4, pp. 362-384.*

Mahmassani H.S. and Chang G.-L. (1985). "Dynamic Aspects of Departure-Time Choice Behavior in a Commuting System: Theoretical Framework and Experimental Analysis". *Transportation Research Record 1037, pp. 88-101.*

Mahmassani H. S. and Chang G.-L. (1986). "Experiments with Departure Time Choice Dynamics of Urban Commuters". *Transportation Research, Vol. 20B, No. 4, pp. 297-320.*

Mahmassani H.S. and Tong C.-C.(1986). "Availability of Information and Dynamics of Departure Time Choice: Experimental Investigation". *Transportation Research Record 1085, pp. 33-47.*

Mahmassani H.S., Chang G.-L. and Herman R. (1986). "Individual Decisions and Collective Effects in a Simulated Traffic Sygstem". *Transportation Science, Vol. 20, No. 4, pp. (258-271).*

Mahmassani H.S. and Chang G.-L. (1987). "On Boundedly Rational User Equilibrium in Transportation Systems". *Transportation Science Vol. 21, No. 2, pp. 89-99.*

Mahmassani H.S. and Herman R. (1987). "Interaction of Trip Decisions and Traffic Systems Dynamics". *European Journal of Operational Research Vol.30, pp. 304-317.*

Matsoukis E.C. (1986). "Road Traffic Assignment - A Review, Part I: Non-Equilibrium Methods". *Transportation Planning and Technology Vol. 11, pp. 69-79.*

Matsoukis E.C. and Michalopoulos P.C. (1986). "Road Traffic Assignment - A Review, Part II: Equilibrium Methods". *Transportation Planning and Technology Vol. 11, pp. 117-135.*

McFadden D. (1976). "The mathematical Theory of Demand Models". In *Behavioral Travel Demand Models*, edited by P.R. Stopher and A.H. Meyburg, Lexington Books, Lexington, Mass.

McFadden D., Tye W. and Train K. (1977). "An Application of the Diagnostic Tests for the Independence from Irrelevant Alternatives Property of the Multinomial Logit Model". *Transportation Research Record 637.*

McFadden D. (1990). "A method of simulated moments for estimation of discrete response models without numerical integration". *Econometrica, Vol. 57, No. 5, pp. 995-1026.*

McShane W. and Crowley K. (1976). "Regularity of Some Detector-Observed Arterial Traffic Volume Characteristics". *Transportation Research Record 596.*

- Merchant D.K. and Nemhauser G.L. (1978a). "A Model and an Algorithm for the Dynamic Traffic Assignment Problem", *Transportation Science Vol. 12, No. 3, pp. 183-199*.
- Merchant D.K. and Nemhauser G.L. (1978b). "Optimality Conditions for a Dynamic Traffic Assignment Model", *Transportation Science Vol. 12, No. 3, pp. 200-207*.
- Minh D.L. (1977). "Discrete Time, Single Server Queue from a Finite Population". *Management Science, Vol. 23, No. 7, pp. 756-767*
- Mirchandani P. and Soroush H. (1987). "Generalized Traffic Equilibrium with Probabilistic Travel Times and Perceptions". *Transportation Science Vol. 21, No. 3, pp.133-152*.
- Moore A. J., Jovanis P.P and Koppelman F.S. (1984). "Modelling the Choice of Work Schedule with Flexible Work Hours". *Transportation Science Vol. 18, No. 2, pp. 141-164*.
- Murchland J.D. (1969). "Road Network Traffic Distribution in Equilibrium". *Operations Research-Verfahren, Vol. 8, pp. 145-183*.
- Nguyen S. (1974). " An Algorithm for the Traffic Assignment Problem". *Transportation Science Vol. 8, No. 3, pp. 203-216*.
- Nguyen S. and Dupuis C. (1984). "An Efficient Method for Computing Traffic Equilibria in Networks with Asymmetric Transportation Costs". *Transportation Science Vol. 18, No. 2, pp. 185-202*.
- Newell G.F. (1971). *Applications of Queueing Theory*, Chapman & Hall, London
- Newell G.F. (1980). "Traffic Flow on Transportation Networks" *MIT Press Cambridge, Mass.*
- Newell G.F. (1987). "The Morning Commute for Nonidentical Travellers". *Transportation Science Vol. 21, No. 2, pp. 74-88*.
- Newell G.F. (1988). "Traffic Flow for the Morning Commute". *Transportation Science Vol. 22, No. 1, pp. 47-58*.
- Ortuzar J. de D. (1983). "Nested Logit Models for Mixed-Mode Travel in Urban Corridors". *Transportation Research Vol. 17A, No. 4*.
- Paine F.T, Nash A.N. Hille S.T. and Brunner G.A. (1976). "Consumer attitudes toward auto versus public transport alternatives". *Journal of Applied Psychology vol. 53 no 6 pp 472-480*.
- Poorzahedy H. (1980). *Efficient Algorithms for Solving the Network Design Problem*, Ph.D. Diss., Northwestern University, Department of Civil Engineering, Evanston, Illinois.
- Poorzahedy H. and Turnquist M.A. (1982). "Approximate Algorithms for the Discrete Network Design Problem". *Transportation Research Vol. 16B, No.1, pp. 45-55*.

Powell W.B and Sheffi Y. (1982). "The Convergence of Equilibrium Algorithms with Predetermined Step Sizes". *Transportation Science Vol 16, No. 1, pp. 45-55.*

Pretty (1988). "Road Pricing: A solution for Hong Kong?". *Transportation Research, Vol. 22A, No. 5, pp. 319-328.*

Rao C.R. (1989). *Linear Statistical Inference and its Applications*, Wiley, New York

Rathi A.K. and Lieberman E.B. (1989). "Effectiveness of traffic restraint for a congested urban network: A simulation study", paper presented at the 68th annual TRB meeting, Jan. 22-26 1989, Washington D.C.

Richardson A. (1982). "Search Models and Choice Set Generation", *Transportation Research Vol. 16A, No. 5/6.*

Richardson A.J. and Taylor M.A.P. (1978). "Travel Time Variability on Commuter Journeys". *Highway Speed Ground Transportation Journal". Vol. 12, No. 1, pp. 77-99.*

Safwat K.N.A and Magnanti T.L. (1988). "A Combined Trip Generation, Trip Distribution, Modal Split, and Trip Assignment Model". *Transportation Science Vol. 18, No. 1, pp. 14-30.*

Safwat K.N.A and Walton C.M. (1988). "Computational Experience with an Application of a Simultaneous Transportation Equilibrium Model to Urban Travel in Austin, Texas". *Transportation Research Vol. 22, No. 6, pp. 457-467.*

Selinger C.S. (1977). "Managing Transportation Demand by Alternative Work Schedule Techniques". *Transportation System Management, Transportation Research Board Special Report 172, National Academy of Sciences, Washington, D.C.*

Sheffi Y. (1979). "A Note on the Turn and Arrival Likelihood Algorithms of Traffic Assignment". *Transportation Research Vol.13B, No. 2, pp. 147-150.*

Sheffi Y. (1985). *Urban Transportation Networks : Equilibrium Analysis with Mathematical Programming Methods*. Prentice-Hall, New Jersey.

Sheffi Y. and Daganzo C.F. (1978a). "Another "Paradox" of Traffic Flow". *Transportation Research Vol 12, No. 1.*

Sheffi Y. and Daganzo C.F. (1978b). "Hypernetworks and Supply-Demand Equilibrium Obtained with Disaggregated Demand Models". *Transportation Research Record No. 673, pp. 113-121.*

Sheffi Y. and Powell W.B. (1981). "A Comparison of Stochastic and Deterministic Traffic Assignment over Congested Networks". *Transportation Research Vol. 15B, No. 1, pp. 53-64.*

Sheffi Y. and Powell W.B. (1982). "An Algorithm for the Equilibrium Assignment Problem with Random Link Travel Times". *Networks Vol. 12, No. 2, pp. 191-207.*

Sheffi Y., Hall R. and Daganzo C.F. (1982). "On the Estimation of the Multinomial Probit Model". *Transportation Research Vol. 16A, No. 5/6.*

Simon H. (1955). "A Behavioral Model of Rational Choice". *Quarterly Journal of Economics, Vol. 69, pp. 99-118.*

- Small K.A. (1982). "The Scheduling of Consume Activities: Work Trips". *The American Economic Review Vol. 72 No. 3, pp. 467-479.*
- Smeed R.J. and Jeffcoate G.O. (1971). "The Variability of Car Journey Times on a Particular Route". *Traffic Engineering and Control Vol. 13, No. 6, pp. 238-243.*
- Smith M.J. (1979a). "Existence, Uniqueness and Stability of Traffic Equilibria". *Transportation Research, Vol 13B, No. 4, pp. 295-304.*
- Smith M.J. (1979b). "The Marginal Cost Taxation of a Transportation Network". *Transportation Research Vol 13B, pp. 237-242.*
- Smith M.J. (1984). "The Existence of a Time-Dependent Equilibrium Distribution of Arrivals at a Single Bottleneck". *Transportation Science Vol 18, pp. 385-394.*
- Sobel K.L. (1980). "Travel Demand Forecasting by Using the Nested Multinomial Logit Model". *Transportation Research Record 775.*
- Swait J, and Ben-Akiva M. (1987). "Incorporating Random Constraints in Discrete Choice Models : An Application to Mode Choice in Sao Paulo", *Transportation Reasearch Vol. 21B, No. 2.*
- Taylor M.A.P. (1977). "Parameter Estimation and Sensitivity of Parameter Values in a Flow-Rate/Travel-Time Relation". *Transportation Science Vol. 11, pp. 275-292.*
- Taylor M.A.P. (1984). "A Note on Using Davidson's Function in Equilibrium Assignment". *Transportation Reasearch Vol. 18B, No. 3, pp. 181-199.*
- Tobin R.L. (1977). "An Extension of Dial's Algorithm Utilizing a Model of Tripmaker's Perceptions." *Transportation Research Vol. 11, No. 5, pp. 337-342.*
- Transporation Research Board (1980). "Alternative Work Schedules: Impacts on Transportation". *National Cooperative Highway Research Program, Synthesis of Highway Practice, 73. National Academy of Sciences, Washington, D.C*
- Transporation Research Board (1985). *Highway Capacity Manual, Special Report 209.* Transportation Research Board, National Research Council, Washington, D.C.
- Van Aerde M. and Yagar S. (1988a). "Dynamic Integrated Freeway/Traffic Signal Network: Problems and Proposed Solutions". *Transportation Research Vol 22A, No. 6, pp. 435-443.*
- Van Aerde M. and Yagar S. (1988b). "Dynamic Integrated Freeway/Traffic Signal Network: A Routing-Based Modelling Approach". *Transportation Research Vol 22A, No. 6, pp. 445-453.*
- Van Knippenberg C. and Van Knippenberg A. (1986). "Discussion". *Transportation Research Record 1085, pp. 47-49.*
- Vickrey W.S. (1969). "Congestion Theory and Transport Investment". *American Economic Review Vol.59, pp. 251-26.*
- Van Vuren T. and Jansen G.R.M. (1988). "Recent Developments in Path Finding Algorithms: A Review". *Transportation Planning and Technology, Vol. 12, pp. 57-71.*



- Vythoulkas P.C. (1988). "Comparison of Dynamic and Static Stochastic Assignment over a Simple Network". *Twentieth Annual UTSG Conference, University College London, Jan. 1988*.
- Vythoulkas P.C. (1989). "A dynamic stochastic assignment model for the analysis of general networks". Paper presented at the U.S.-Italy joint seminar on Urban Traffic Networks: Dynamic Control and Flow Equilibrium. *Transportation Research, Vol. 24B, No. 6, pp. 453-469*.
- Vythoulkas P.C. (1990a). "Two Models for Predicting Dynamic Stochastic Equilibria in Urban Transportation Networks". *Proceedings of the 11th International Symposium on Transportation and Traffic Theory, held at Yokohama, July 1990. M. Koshi ed., Elsevier, New York, 1990, pp. 253-272*.
- Vythoulkas P.C. (1990b). "Evaluating the effects of Road Pricing", Working paper No. 611, Transport Studies Unit, University of Oxford.
- Wardrop J.G. (1952). "Some Theoretical Aspects of Road Traffic Research". *Proceedings, Institution of Civil Engineers 2a, pp.325-378*.
- Watson P.L. and Holland E.P. (1978). "Relieving Traffic Congestion: The Singapore Area License Scheme". *World Bank Staff Working Paper No. 281, World Bank, Washington, D.C.*
- Weintraub A. and Ortiz C. and Gonzalez (1985). "Accelerating Convergence of the Frank-Wolfe Algorithm". *Transportation Research, Vol. 19, No. 2, pp. 113-122*.
- Williams H.C.W.L. (1977). "On the Formation of Travel Demand Models and Economic Evaluation Measures of User Benefit". *Environment and Planning A, Vol 9, pp. 285-344*.
- Wilde D.J. (1964). *Optimum Seeking Methods*. Prentice-Hall, Englewood Cliffs, New Jersey.
- Wohl M. (1970). "A Methodology for Forecasting Peak and Off-Peak Travel Volumes". *Highway Research Record 322, pp. 182-219*.
- Yagar S. (1971). "Dynamic Traffic Assignment by Individual Path Minimisation and Queueing". *Transportation Research, Vol. 5, No. 5, pp. 176-196*.
- Yagar S. (1976). "Emulation of Dynamic Equilibrium in Traffic Networks". *Lecture Notes in Economics and Mathematical Systems, 118 : Traffic Equilibrium Methods, Montreal 1974., Ed. M. Florian, Springer-Verlag, Berlin, pp. 240-264*.
- Zangwill W.I. (1969). *Nonlinear Programming: A Unified Approach*. Prentice-Hall, Englewood Cliffs, N.J.
- Zipkin P.H. (1980). "Bounds for Aggregating Nodes in Network Problems". *Mathematical Programming, Vol 19, pp. 155-177*.
- Zawack D.J. and Thompson G.L. (1987). "A Dynamic Space-Time Network Flow Model for City Traffic Congestion". *Transportation Science Vol. 21, No. 3, pp. 153-162*.

# A1 The Gumbel distribution : basic properties

Assume that  $x$  is Gumbel distributed. Then the cumulative distribution function of  $x$  is given by :

$$F(x) = \exp[-e^{-\mu(x-n)}], \quad \mu > 0$$

and the probability density function by :

$$f(x) = \mu \cdot e^{-\mu(x-n)} \exp[-e^{-\mu(x-n)}]$$

where  $n$  is a locational parameter and  $\mu$  is a positive scale parameter.

This distribution has the following properties :

1. The mode is  $n$
2. The mean is  $n + \gamma/\mu$ , where  $\gamma$  is Euler constant ( $\sim 0.577$ )
3. The variance is  $\pi^2/6\mu^2$ .
4. The Gumbel distribution is preserved over linear transformations. If  $x$  is Gumbel distributed with parameters  $(n, \mu)$ , and  $\beta$  and  $a > 0$  are any scalar constants, then  $a \cdot x + \beta$  is Gumbel distributed with parameters  $(a \cdot n + \beta, \mu/a)$
5. If  $x_1$  and  $x_2$  are independent Gumbel-distributed variates with parameters  $(n_1, \mu)$  and  $(n_2, \mu)$ , respectively, then  $x^* = x_1 - x_2$  is logistically distributed :

$$F(x^*) = \frac{1}{1 + e^{\mu(n_2 - n_1 - x^*)}}$$

6. If  $x_1$  and  $x_2$  are independent Gumbel distributed with parameters  $(n_1, \mu)$  and  $(n_2, \mu)$ , respectively, then

$$\max(x_1, x_2)$$

is Gumbel distributed with parameters

$$\left( \frac{1}{\mu} \ln(e^{\mu n_1} + e^{\mu n_2}), \mu \right)$$

7. As a corollary to proposition 6, if  $(x_1, x_2, \dots, x_J)$  are  $J$  independent Gumbel-distributed random variables with parameters  $(n_1, \mu), (n_2, \mu), \dots, (n_J, \mu)$ , respectively, then  $\max(x_1, x_2, \dots, x_J)$  is Gumbel distributed with parameters

$$\left( \frac{1}{\mu} \ln \sum_{j=1}^J e^{\mu n_j}, \mu \right)$$

# A2 Network data

LINK	CAPACITY (veh/h)	FREE FLOW SPEED (km/h)	LENGTH (km)	LINK	CAPACITY (veh/h)	FREE FLOW SPEED (km/h)	LENGTH (km)	LINK	CAPACITY (veh/h)	FREE FLOW SPEED (km/h)	LENGTH (km)
1	3800.	60.0	6.0	25	2400.	55.0	3.0	49	2600.	45.0	1.5
2	3200.	45.0	1.5	26	2600.	50.0	3.5	50	2800.	45.0	3.5
3	3400.	50.0	4.0	27	2700.	45.0	2.0	51	2200.	45.0	2.5
4	3000.	45.0	2.0	28	2400.	45.0	3.0	52	2500.	45.0	3.0
5	3200.	50.0	4.0	29	2800.	50.0	1.5	53	3200.	45.0	2.5
6	3400.	55.0	2.0	30	3800.	60.0	6.0	54	2700.	45.0	1.5
7	3000.	45.0	2.5	31	3400.	55.0	2.0	55	2200.	45.0	2.5
8	2700.	45.0	2.0	32	3200.	45.0	1.5	56	2700.	45.0	3.5
9	2800.	50.0	4.5	33	3400.	50.0	4.0	57	3000.	50.0	3.5
10	2400.	45.0	2.5	34	3600.	55.0	5.0	58	2700.	45.0	3.5
11	3800.	60.0	6.0	35	3000.	45.0	2.0	59	2400.	45.0	2.5
12	3600.	55.0	5.0	36	3000.	45.0	2.5	60	3000.	45.0	3.0
13	3400.	50.0	2.5	37	2800.	45.0	1.5	61	2700.	45.0	2.0
14	3800.	60.0	6.0	38	2800.	45.0	1.5	62	2500.	45.0	3.0
15	3400.	50.0	2.0	39	2600.	45.0	2.0	63	2400.	45.0	2.5
16	2800.	50.0	3.0	40	2600.	45.0	1.5	64	2400.	45.0	2.0
17	3000.	45.0	3.0	41	2700.	45.0	2.0	65	2400.	45.0	3.0
18	2400.	45.0	2.0	42	2400.	45.0	2.5	66	2400.	45.0	1.5
19	2800.	50.0	4.5	43	2600.	45.0	2.0	67	3400.	50.0	2.0
20	3200.	60.0	6.0	44	2700.	45.0	1.5	68	2400.	45.0	1.5
21	3200.	45.0	2.5	45	3200.	45.0	2.5	69	3000.	55.0	3.0
22	3200.	60.0	6.0	46	3400.	45.0	3.5	70	2400.	55.0	3.0
23	2600.	50.0	3.5	47	3800.	45.0	3.5	71	2800.	50.0	1.5
24	3000.	50.0	3.5	48	2800.	45.0	3.5	72	3000.	55.0	3.0