

**ESTIMATION OF TRIP MATRICES FROM TRAFFIC COUNTS:
AN EQUILIBRIUM APPROACH**

by

JAE HAK OH

A thesis submitted to the University of London
for the degree of Doctor of Philosophy

Transport Studies Group
University College London

March 1991

ABSTRACT

In urban traffic management and planning, an important problem is how to obtain estimates of origin-destination (O-D) trip matrices using low-cost data such as traffic counts. Although conventional methods using the data from direct surveys can be used to estimate trip matrices, they appear to be inaccurate and expensive. By contrast, the use of traffic counts is attractive, as it is less expensive and more practical.

The main objective of the research reported in this thesis is to develop new methods for estimating trip matrices from traffic counts when congestion effects in networks are considered. The problem and existing methods including the sequential solution method used in the ME2 model are reviewed.

A new formulation is given for the problem which solves the two sub-problems of entropy maximization and equilibrium traffic assignment simultaneously. It allows modelled link flows to be constrained so as to equal observed ones without link assignment proportions of the trips. A simultaneous solution method is presented for this new formulation. To reduce the considerable computational burden in solving the problem, a heuristic method has been developed which uses a linear approximation fitted by regression to the equilibrium link flows. Extrapolation and perturbation methods have also been used to speed up the solution process. However, the simultaneous solution method appears to be impractical for use in large networks because of the heavy computational demand. As an alternative, an improved sequential solution method is proposed which uses a penalty function method. This method approximates a solution by solving a sequence of problems, while fixed link assignment proportions are used.

The performance of the proposed methods has been tested and compared with that of the existing sequential ME2 method using both small example networks and a larger real network. The results show that the simultaneous method works well and that it performs better than the

existing sequential method or the improved sequential method. The improved sequential method is also shown to perform closely to the simultaneous one. Some practical implications of the new methods including the robustness of the solutions and the increased computational burden are discussed and they are also compared with those of the sequential solution method.

The conclusions from the main findings of the research are drawn and a number of suggestions for further study are given.

ACKNOWLEDGEMENTS

I would like to thank Dr Benjamin G. Heydecker for his patient supervision and encouragement throughout the period of this research. The completion of this thesis would be difficult without his steady advice.

I wish to express my gratitude to Professor Richard Allsop and Dr Luis Willsumsen for their useful discussions and advice from time to time.

Many thanks are also due to the members of the Transport Studies Group at UCL for their help and warm friendship throughout the period of this research.

This research has been carried out with the financial support of the British Council. I would like to thank them for their support.

I owe a debt of gratitude to our families in Korea for their warm consideration and supporting. In particular, I would like to thank my father-in-law for his financial support.

Finally, but most importantly, I would like to thank my wife, Seonju, for her patient supporting and understanding, and I would also like to share my pleasure with our loving daughter, Keunyoung. It would not be possible to complete this thesis without their help.

CONTENTS

	Page
ABSTRACT	2
ACKNOWLEDGEMENTS	4
CONTENTS	5
LIST OF TABLES	8
LIST OF FIGURES	10
CHAPTER 1. INTRODUCTION	
1.1 Background	12
1.2 Objectives	14
1.3 Structure of the thesis	14
CHAPTER 2. ESTIMATION OF TRIP MATRICES	
2.1 Introduction	15
2.2 Modelling of study areas and trip matrices	15
2.3 Conventional methods for estimating trip matrices	19
2.3.1 Direct estimation methods	19
2.3.2 Indirect estimation methods	21
2.3.3 Errors in trip matrices estimated by conventional methods	24
2.4 Simplified methods using traffic counts	26
CHAPTER 3. THE ESTIMATION PROBLEM USING TRAFFIC COUNTS	
3.1 Introduction	29
3.2 Review of traffic assignment methods	29
3.2.1 Proportional traffic assignment methods	32
3.2.2 Equilibrium traffic assignment methods	34
3.3 The estimation problem	39
3.3.1 Description of the problem	39
3.3.2 Treatment of underspecification	42
3.3.3 Treatment of inconsistent traffic counts	42
3.3.4 Treatment of traffic congestion effects	44

CHAPTER 4. REVIEW OF RELATED METHODS

4.1 Introduction	47
4.2 Classification of existing estimation methods	48
4.3 Methods using proportional traffic assignment	50
4.3.1 Methods for calibrating demand models	51
4.3.2 Methods based on information theory	57
4.3.3 Methods using statistical inference techniques	72
4.4 Methods using equilibrium traffic assignment	84
4.4.1 Willumsen's method	84
4.4.2 Nguyen's method	94
4.4.3 Fisk and Boyce's method	102

CHAPTER 5. SIMULTANEOUS ESTIMATION OF TRIP MATRICES

5.1 Introduction	106
5.2 Formulation	108
5.2.1 Entropy objective functions	108
5.2.2 Link flow constraints	111
5.2.3 A new formulation	112
5.3 A simultaneous solution method	112
5.3.1 Use of sequential unconstrained optimization methods	113
5.3.2 Linear approximation to equilibrium link flows	116
5.3.3 A solution algorithm	117
5.3.4 Use of an extrapolation method	121
5.3.5 Use of a perturbation method	123
5.3.6 Interface with an equilibrium assignment program	125
5.3.7 Computational demand of the simultaneous solution method	126
5.4 An improved sequential solution method	126

CHAPTER 6. TESTS OF THE ESTIMATION METHODS

6.1 Introduction	131
6.2 Tests with a simple network	132
6.2.1 Introduction	132
6.2.2 Results and comments	137
6.3 Tests with the Nguyen's network	141
6.3.1 Introduction	141
6.3.2 Objectives and design of tests	142

6.3.3 Base test	145
6.3.4 Tests with various congestion levels	158
6.3.5 Tests with inconsistent link flows	163
6.3.6 Tests with determinate matrix total	168
6.3.7 Tests with different prior trip matrices	170
6.3.8 Tests with different sets of links observed	173
6.4 Tests with the Ripon network	178
6.4.1 Introduction	178
6.4.2 Tests with assigned link flows	183
6.4.3 Tests with real link flows	198

**CHAPTER 7. CONCLUSIONS AND SUGGESTIONS
FOR FURTHER STUDY**

7.1 Conclusions	211
7.2 Suggestions for further study	214

REFERENCES	216
-------------------	-----

APPENDIX 1: LIST OF MAJOR NOTATION	229
---	-----

APPENDIX 2: NGUYEN'S EXAMPLE NETWORK DATA	232
--	-----

LIST OF TABLES

	Page
4.1 Comparison of observed trip matrices and link flows on different days (taken from Van Vliet and Willumsen (1981))	63
4.2 Comparison of observed vs estimated trip matrices using known route choice and all or nothing assignment on Tuesday (taken from Van Vliet and Willumsen (1981))	64
4.3 Comparison of the observed and modelled flows by the mean absolute difference divided by the average observed flow expressed as a percentage (taken from Matzoros et al (1987))	65
4.4 Estimating a trip matrix using a simple example by the sequential solution method	86
4.5 Observed vs loaded flows from 19 October trip matrices (extracted from Willumsen (1981a))	90
4.6 Tests with the sequential method for estimating trip matrices (extracted from Willumsen (1981a))	91
6.1 Results of the tests using the simple example network	138
6.2 Results of the base test by ME2	146
6.3 Results of the base test by SIM-S ₀	148
6.4 Results of the base test by SIM-S ₁	148
6.5 Results of the base test by ME3-S ₀	152
6.6 Results of the base test by ME3-S ₁	152
6.7 Summary of the results of the base test by various methods	156
6.8 Results of the tests with various congestion levels	159
6.9 Results of the tests with inconsistent link flows	166
6.10 Results of the tests with determinate matrix totals	169
6.11 Results of the tests with different prior trip matrices	171
6.12 Results of the tests with different sets of links observed	176
6.13 Results of the test using assigned link flows by ME2	184
6.14 Results of the test using assigned link flows by SIM-S ₁	186
6.15 Results of the test using assigned link flows by SIM-S ₀	188
6.16 Results of the test using assigned link flows by ME3-S ₁	190
6.17 Results of the test using assigned link flows by ME3-S ₀	190
6.18 Computing times incurred to the alternative solution methods	197
6.19 Results of the test using real link flows by ME2	199
6.20 Results of the test using real link flows by SIM-S ₁	201

6.21 Results of the test using real link flows by SIM-S0	203
6.22 Results of the test using real link flows by ME3-S1	205
6.23 Results of the test using real link flows by ME3-S0	205

LIST OF FIGURES

	Page
3.1 Simple network showing an underspecified problem	41
3.2 Simple network showing variable assignment proportions	45
4.1 A simple example network used for testing of the convergence of the sequential method	85
4.2 Estimating a trip matrix by the sequential ME2 method in the SATURN (taken from Van Vliet (1987))	92
5.1 Flow chart of the simultaneous solution method (Algorithm A5.1)	119
5.2 Flow chart of the improved sequential solution method (Algorithm A5.4)	129
6.1 The simple example network	132
6.2 The speed-flow curve for the simple example network	133
6.3 Nguyen's (1977) example network	142
6.4 Results of the base test by ME2	147
6.5 Results of the base test by SIM-S ₀	149
6.6 Results of the base test by SIM-S ₁	150
6.7 Results of the base test by ME3-S ₀	153
6.8 Results of the base test by ME3-S ₁	154
6.9 Summary of the results of the base test by various methods	156
6.10 Results of the tests with various congestion levels	161
6.11 Results of the tests with inconsistent link flows	164
6.12 Results of the tests with different sets of links observed	174
6.13 Map of the town of Ripon	179
6.14 Ripon network	180
6.15 Results of the test using assigned link flows by ME2	185
6.16 Results of the test using assigned link flows by SIM-S ₁	187
6.17 Results of the test using assigned link flows by SIM-S ₀	189
6.18 Results of the test using assigned link flows by ME3-S ₁	191
6.19 Results of the test using assigned link flows by ME3-S ₀	192
6.20 Comparison of the trade-off curves estimated from assigned link flows by various solution methods	194
6.21 Comparisons of goodness of fit of the trip matrices estimated by various solution methods	195

6.22	Comparisons of goodness of fit vs gap, goodness of fit vs entropy(S_1), and goodness of fit vs entropy(S_0)	196
6.23	Results of the test using real link flows by ME2	200
6.24	Results of the test using real link flows by SIM- S_1	202
6.25	Results of the test using real link flows by SIM- S_0	204
6.26	Results of the test using real link flows by ME3- S_1	206
6.27	Results of the test using real link flows by ME3- S_0	207
6.28	Comparison of the trade-off curves from real link flows by various solution methods	208

CHAPTER 1. INTRODUCTION

1.1 Background

The estimation of origin-destination (O-D) trip matrices is an important part in the analysis of traffic management and transport planning tasks. The trip matrices are often used to design and evaluate the new transport plans. For example, to assess the impacts of alternative traffic management schemes, one makes use of the trip matrices as an input to traffic assignment methods in order to estimate the likely variations in link flows on the network.

Conventional methods using the data from direct surveys can be used to estimate trip matrices. However, they are fairly expensive, involving considerable resources and causing interruptions to trip makers. More importantly, their end products may well be short-lived and unreliable. All methods of this kind use the sampled data and thus they can only provide an approximation of the trip matrix for the survey period.

By contrast, the use of traffic counts for estimating trip matrices can avoid at least some of the difficulties identified in conventional methods. In particular, traffic counts are easily available or inexpensive to collect. Traffic counts are collected regularly by local authorities for various local traffic management uses. Most counting operations can be performed without interrupting traffic or causing delays to travellers. Furthermore, traffic counts are automatically collected from the detectors laid down under the road. Secondly, as traffic counts can be collected routinely, the evolution of the data base can be easily followed up. This enables transport planners to update designs and forecasts continuously. Finally, the new methods using traffic counts are simple in terms of data processing and the estimation process. However, they could produce more accurate results, since they use more reliable data.

Since the potential of using traffic counts for estimating trip matrices was recognized, various methods have been developed. The most

common idea of estimating trip matrices from traffic counts is to find a trip matrix which, when assigned to the network, closely replicates traffic counts. Accordingly, the use of an appropriate traffic assignment method is important in the matrix estimation process. According to the review of the literature presented in Chapter 4, the problem of estimating a trip matrix from traffic counts when the route choice proportions from the traffic assignment are assumed to be fixed over variation of traffic demands is now well researched. However, the use of proportional assignment methods such as all or nothing assignment is not sufficiently realistic, especially when congestion in networks plays an important role in route choice. The better result in the matrix estimation of using traffic counts can be achieved by using more advanced assignment methods such as Wardrop's equilibrium traffic assignment.

A number of methods have been proposed for the problem of estimating trip matrices from traffic counts when the traffic equilibrium conditions are taken into account. They can be classified into three methods: Willumsen's method (Hall, Van Vliet and Willumsen, 1980), Nguyen's method (Nguyen, 1977) and Fisk and Boyce's method (1983). However, none of these three methods solves the problem satisfactorily. In particular, Willumsen's method appears to be attractive because of its advantages such as the simple data requirement and the low computing cost. The method was initially developed by assuming fixed route choice proportions for the trips and was later extended to use equilibrium assignment. However, the sequential solution method applied to solve the extended problem is only a heuristic, as it solves the two subproblems of entropy maximization and equilibrium assignment alternately. The sequential solution method cannot be guaranteed to converge to optimal solutions or even to converge at all. On the other hand, Nguyen's method has the form of a traffic assignment problem with elastic demand and it uses a set of the interzonal travel costs as the input data which may be obtained from traffic counts. Fisk and Boyce's method is an extension of a doubly constrained gravity model whose applications may not be suitable for urban transport studies. These two methods are distinguished from Willumsen's one in a sense that they are based on the different level of the detail in the input data.

1.2 Objectives

The main objective of this research is to develop new methods for estimating trip matrices from traffic counts when congestion effects in networks are considered. The problem and existing methods for estimating trip matrices from traffic counts are reviewed. A new formulation and its solution methods are proposed and their performance is tested and compared with that of the existing sequential solution method.

1.3 Structure of the thesis

This thesis is organized as follows. **Chapter 1** describes a general background of this study with specific objectives and outlines the structure of the thesis. **Chapter 2** provides an overview of some background to trip matrix estimation. Various conventional methods for estimating trip matrices are reviewed and their short-comings are discussed. As an alternative to conventional methods, the advantages and disadvantages of the simplified methods for estimating trip matrices from traffic counts are addressed.

Chapter 3 identifies some fundamental issues in the formulation of the estimation problem using traffic counts. Two main types of traffic assignment methods - proportional assignment and capacity restrained assignment - are reviewed. **Chapter 4** provides a detailed and up-to-date review of various existing methods for estimating trip matrices from traffic counts.

Chapter 5 proposes a new formulation and solution methods to estimate a trip matrix from traffic counts under equilibrium traffic conditions. The simultaneous method is shown to be impractical in terms of computing time for large networks. As an alternative, an improved sequential method is proposed. **Chapter 6** tests the proposed estimation methods. Three test cases are investigated and their results are reported in detail.

Finally, **Chapter 7** draws the conclusions from the main findings of this study and a number of suggestions for further study are also given.

CHAPTER 2. ESTIMATION OF TRIP MATRICES

2.1 Introduction

This chapter provides an overview of some basic background knowledge on trip matrix estimation. The estimation of trip matrices is an important step and also an expensive stage involving extensive resources in the transportation planning process. During the past decades, a large amount of effort by many researchers and practitioners has been devoted to develop methods of providing more accurate, more consistent and less expensive end products. As a result, a number of the methods have been developed. It is not possible and unnecessary for this research to investigate the details of all of those methods. This chapter reviews only some of the methods which are considered to be most important and relevant to this research.

This chapter is organised as follows. Section 2.2 makes a definition of a transport planning network, trip matrices, and their mathematical notations conventionally used in transportation planning studies. Section 2.3 reviews some of conventional methods for trip matrix estimation with a particular attention to their major shortcomings. Section 2.4 discusses leading motivations of developing simplified methods apart from complicated conventional methods. As an alternative to conventional methods, the method of estimating a trip matrix from traffic counts is assessed briefly and compared to conventional methods.

2.2 Modelling of study areas and trip matrices

A study area is modelled in transportation planning studies by a zoning system. A zoning system consists of zones and zone centroids. Each zone represents a sub-area of the study area, where each zone has a centroid as a center to represent activities within the zone. All trips are assumed to originate from zone centroids and terminate there. A study area can also be considered as divided into two areas: an internal one representing the area of interest itself, and an external one

representing the rest of the transport system in so far as it affects the internal part. The internal area is expressed by internal zones and the external one by external zones.

A road network is expressed by a set of N nodes and a set of L links. In this study, emphasis is placed on the consideration of typical urban road networks. Nodes are usually associated with points of interest in the network such as junctions, and they are labelled consecutively from $n=1$ to N . A link is represented by an ordered pair of nodes, if there is a link between two nodes. Links are always one way and for some purposes it is useful to associate successive numbers to them, $a=1,2,\dots,L$. Each link is associated with a number of attributes. The most important attributes include:

- link length: physical distance along the road, d_a , measured in meters,
- design speed: average free flow speed, s_a , measured in km/hr,
- link capacity: maximum rate of making traffic passing, Q_a , measured in vehs/h. It depends on factors such as number of lanes or width,
- travel time, t_a , measured in minutes,
- travel cost, c_a , usually a weighted combination of travel time and distance,
- flow or traffic volume, V_a , measured in pcu/h or vehs/h,
- a cost-flow relationship, $c_a(V_a)$, a function of the amount of traffic using link 'a' relating travel cost on the link to link volumes.

Centroid connectors are used to connect zone centroids with nodes in the network. The cost associated with their use represents the average cost of travelling over the local streets from the origin before

joining the main network. This cost is normally considered to be independent of the traffic flow.

A trip between origin zone i and destination zone j will use a particular sequence of links called a path or route and T_{ijr} will be the trips from i to j which use route r . The cost of travelling along this route is the sum of the costs of the individual links used and will be represented by C_{ijr} . The variable δ_{ijr}^a can be used to identify links used by route r between origin i and destination j . Then, the route cost is

$$C_{ijr} = \sum_a c_a \delta_{ijr}^a \quad (2.1)$$

where $\delta_{ijr}^a = 1$ if link 'a' is used by route r between i and j , and otherwise, $\delta_{ijr}^a = 0$.

Preparation of study networks lays a basis for transportation planning studies. The study area should be defined with a greater care. The most important criterion is that it should include those routes which would be affected significantly by traffic rerouting as a result of any proposal. It is desirable that study area boundaries should coincide with administrative boundaries to facilitate the provision of data. After defining the study area, the zoning system and the road network in the study area will be modelled depending on the size of the study area and the type of the study. For example, large-scale strategic planning studies may be based on coarse networks, which mean a higher level of aggregation. Small scale traffic management studies require finer networks, which means a low level of aggregation. The definition of the study area, zoning system and road network is a compromise between conflicting requirements between the desire to improve the model by increasing its size and complexity and the practical considerations of keeping costs down and meeting targets of time scale and adequacy. Experience and professional judgment will often be important components. For more useful disciplines for defining study areas and modelling networks, see 'Traffic Appraisal Manual (DTp, 1981)' and 'Road and Traffic in Urban Areas (IHT and DTp, 1987)'

The number of trips per unit time from origin i to destination j is represented by T_{ij} and the complete set of trips covering all the study area constitutes the trip matrix $\underline{T}=\{T_{ij}\}$. Thus, a trip matrix is a representation of the trips made between pairs of zones in the study area.

Trip matrices are an important element in the analysis of traffic management and transport planning tasks. Trip matrices are often used to design and evaluate the new transport plans. For example, to assess the impacts of alternative traffic management schemes, one makes use of the trip matrices as an input to traffic assignment methods in order to determine variations in link flows on the network.

The trip matrix depends on two forms of aggregation, spatial and temporal. Spatial aggregation involves the grouping of areas into discrete spatial units or zones. In large-scale modelling exercises, the number of zones can be thousands, whereas for small scale traffic management schemes 25 to 50 zones may be enough. But even in this latter case, the number of cells is quite large and many of them are likely to contain zeros or small numbers of the order of 1 trip per hour.

Temporal aggregation is concerned with the time interval during which trips between zones are considered. The choice of this time slice or interval has a major impact on the trip matrix and its usefulness in the analysis of particular problems. In a detailed analysis of a system of saturated traffic signals, time slices of the order of 15 minutes are used to follow the build-up and decay of queue lengths and travel times. On the other hand, most traffic management problems require discrimination at the hourly level only and many problems involving new road construction could be handled with trip matrices based on 16 or 24 hours. Indeed, some analyses such as inter-city demand studies, use weekly, monthly or yearly trip matrices.

Trip matrices based on a small time slice present some particular problems. For a start, it may be difficult to allocate unequivocally trips to time slices. For example, consider the difficulty of allocating a trip starting at 9:10 a.m. and ending at 9:20 a.m. in the study area

to either the 9:00 to 9:15 or 9:15 to 9:30 trip matrix. Thus, it is desirable to use time slices which are considerably longer than a typical trip length in the study area. The second problem is sparsity: the smaller the time slice, the greater will be the number of cells in the trip matrix which contains fewer than 1 trip.

The objectives of a particular study requiring O-D information will help to define the spatial and temporal aggregation for the trip matrix used.

Trip matrices are subject to hourly, daily, weekly and seasonal variations in the same way as traffic counts or trip rates. One can think of a distribution of trip matrices over, say, a year and it will depend again on the objectives and resources of the study which matrix can be said to be the trip matrix of interest.

2.3 Conventional methods for estimating trip matrices

This section provides an overview of conventional methods for estimating trip matrices. For a practical purpose, conventional methods are classified into two main types: direct estimation methods such as roadside interviews which use direct measurements of trip matrices and indirect estimation methods or synthetic methods such as distribution models which use other data to infer a trip matrix.

2.3.1 Direct estimation methods

Direct estimation methods rely on field observations, extensive surveys and interviews. The disruptive nature of interviews inevitably causes inconveniences to the travellers and delay to the traffic. Time and budget constraints also limit the number of interviews that can be conducted. Therefore, errors will be inevitable wherever 100 per cent sampling cannot be achieved. Moreover, observation methods rely heavily on the reliability of interviewers and observers in data collection and interpretation. Hence, technical and human errors are likely to occur, e.g. miscoding and misinterpretation of information. Perhaps, the most influential problem associated with this approach is the requirement of

large amount of resources. Willumsen (1981a) provides a good review for various survey methods. More practical details are given in the 'Traffic Appraisal Manual (DTp, 1981)'. Major techniques frequently used are as follows:

(1) **Home interview:** This method is employed in conventional transport planning studies for large towns, major conurbations and regions. It is fairly expensive technique, usually involving large number of staff carrying out interviews in a sample of households or work places. The data thus collected covers a wide range of information about the origin and destination of trips which is given a good deal of attentions. Because of the large cost of collecting and processing home interview data, only a sample of all households is surveyed. Sampling rates normally range between 1 and 10 per cent. As information of different sorts and for a variety of purposes is collected, the sampling rates are a compromise between these objectives and survey costs.

(2) **Roadside interview:** This method requires vehicles to be stopped and questioned regarding the origin and destination of their journeys and other data. These interviews are usually taken on the roads of a cordon at screen line points. Interview stations should be selected in such a manner that all relevant inter-zonal traffic can be sampled. This requires a careful definition of the zones and network in the study area. The number of interviews to be made at each survey station is defined according to sample considerations. Based on sample size and survey period, correction factors are applied in order to expand the original survey data to estimate a trip matrix for a different period. Sampling fractions of 1 in 5 may be considered typical but they certainly depend on traffic levels and manpower available. These interviews tend to be expensive in manpower, delay and disruption to trip makers.

(3) **Flagging methods:** This method requires observers located at each of the entry and exit points of a study area and the use of types of flags to identify vehicles. Colored or numbered stickers are attached on entry and recorded at intermediate and exit points. For small study areas, it is possible to identify vehicles by asking drivers at one entry point to

switch their headlights on for a fixed period of time. Observers at key points then record the number of vehicles with their lights on for given intervals of time. The process is then repeated for different entry points on successive days. This method can only be used during daylight and for small study areas.

(4) **Vehicle following method:** The method requires observers to follow vehicles through the study area recording its passage through key points in the network. This method seems more appropriate for route choice than O-D studies and probably is only advantageous in large and busy central areas.

(5) **Aerial photography:** This method is based on time-lapse aerial photography of a study area from a stationary helicopter hovering at a fixed altitude. The data collection stage is fast and inexpensive compared with alternative methods but this is achieved at the expense of processing effort. The method requires following individual vehicles frame by frame through the study area. This method will be more attractive when automatic identification of vehicles in the frames is available. In principle, a sampling ratio of 1 to 1 for the survey period could be achieved with this method, but practical reasons restrict sampling ratios to values similar to road side interviews.

All the direct estimation methods described above are fairly expensive involving considerable resources and more importantly the end products may well be short-lived and unreliable. Moreover, roadside interview method causes interruptions and delay to traffic. Also, all the methods imply sampling and thus they can only provide an estimate of the trip matrix for the survey period.

2.3.2 Indirect estimation methods

Trip matrices often cannot be obtained directly from surveys mentioned above (e.g., because an excessively large survey is required or suitable survey locations are not available). Alternative techniques are developed to derive a trip matrix from incomplete survey data by means of synthetic models. First, the conventional four stage approach

is briefly described. Then, the two most frequently used distribution models in urban transportation planning - the doubly-constrained gravity model and the partial matrix technique - are discussed.

The conventional four-stage approach used in most transportation planning studies consists of variants of the following sequence.

(1) **Trip generation and attraction:** The number of trips originating from and attracted to each zone is estimated. The generation and attraction for each zone are estimated by trip end models derived using planning data of the existing trip rates and the socio-economic, land use and transport system characteristics of the study area. This is an expensive process involving large volume of survey data.

(2) **Trip distribution:** The number of trips between each origin and destination is estimated using the trip ends generated in the trip generation stage.

(3) **Modal split:** This stage models the choice of mode for trips made, usually car and one or more public transport mode. It is particularly required in planning public transport services.

(4) **Trip assignment:** This stage allocates the trips between zones to various routes that are most likely to be taken in travelling between each pair of zones. The final output is a set of link flows and travel costs between each pair of zones. Various methods for assigning trips into the network will be discussed in greater detail in Chapter 3.

There are many other variations for the trip distribution models. Here only the following two distribution models are discussed.

(1) **The gravity model:** The most widely used gravity model is the doubly-constrained gravity model (Wilson, 1967).

$$T_{ij} = A_i O_i B_j D_j f(C_{ij}) \quad (2.2)$$

where O_i and D_j are total number of trips generated and attracted to zones i and j , A_i and B_j are balancing factors calculated as

$$A_i = \frac{1}{\sum_j B_j D_j f(C_{ij})} \quad (2.3)$$

$$B_j = \frac{1}{\sum_i A_i O_i f(C_{ij})} \quad (2.4)$$

and $f(C_{ij})$ is a deterrence function or measure of separation with at least one parameter for calibration. The most popular form of the deterrence function is the exponential type, $f(C_{ij}) = \exp(-\beta C_{ij})$ with the cost perception parameter β . The parameters of the deterrence function are usually calibrated so that the model produces a trip length distribution which is as close as possible to that obtained from survey data.

In principle, the parameters may not be transferable between study areas, since they are calibrated for each particular set of data. The gravity model is simple to calibrate, but it may have limitations in the representation of real trip-making behavior.

(2) Partial matrix technique: Partial matrix techniques have been developed in order to synthesize a trip matrix using incomplete data. It is used to fill in missing cells of a trip matrix. Completed cells are used to derive a relationship between interzonal trips, zonal characteristics and interzonal travel costs from which missing cells are generated. The theoretical basis for this approach was put forward by Kirby (1979) and it has been practically applied by Neffendorf and Wootton (1974) and many others. The technique is attractive because of its survey cost saving potential, but many questions remain regarding the errors involved and the choice of good survey patterns (Kirby, 1979).

A main feature of the conventional four stage approach is that it is sequential, as each stage is based on the output from the preceding stage. There might be internal inconsistencies within this sequence. For

example, the trip generation stage - the estimation of trip ends - assumes implicitly a general level of service in the network. The trip distribution stage requires an explicit level of service such as travel costs in the network. The explicit level of service is calculated from the trip assignment stage. However, the initial levels of service used for trip generation and distribution are rarely revised to be consistent with the travel costs calculated by the trip assignment stage. This approach requires extensive input data for trip generation. The input data is not only expensive to collect but also they are likely to be unreliable and inaccurate.

2.3.3 Errors in trip matrices estimated by conventional methods

Any trip matrix, whether obtained from direct surveys or indirect transport modelling exercises will only be an approximate representation of the actual trip matrix. Thus, the resulting trip matrix will not be free from errors. It is important to understand the errors and inaccuracies in the trip matrices estimated by conventional methods described in Sections 2.3.1 and 2.3.2. Willumsen (1981a) mentioned that the accuracy of a trip matrix obtained by the direct and indirect estimation methods are subject to a range of sources of errors as follows:

(1) Daily/seasonal variations and survey period expansion errors: This type of errors occurs when correction factors are applied to convert the original survey data to get a different time slice or period trip matrix (for example to expand a 16 hour survey to 24 hour survey). It may well be the case if a trip matrix is to be produced for a longer time period but survey data is only available for a shorter time slice. These errors are mainly caused by the time variations of the trip matrix.

(2) Data collection errors: This type of errors occurs during the survey period and is mainly caused by human errors, for example, misreporting of trips, misidentification of vehicles, incomplete questionnaires or even errors while writing down information. This type of error is usually encountered and practically unavoidable. Good quality control can decrease but not eliminate these errors.

(3) **Data processing errors:** This type of errors occurs in the process of transferring and compiling the raw data set and is mainly caused by human errors. Miscoding, incorrect typing of data, double counting, missing records, editing a checking list and creation of files, production of output tables and even programming errors are some of the main sources. Again a good quality control system may help to decrease these errors.

(4) **Sampling errors:** Traffic models are usually based on sampled data, which is taken to be representative of the population concerned. Provided the sampling frames are understood, confidence levels can be calculated for the values being estimated. This type of error occurs because, except in very simple cases, the survey cannot cover all the trips during the survey period. This may be due to the location of the survey sites in roadside interviews or flagging methods which makes it impossible to sample all the trips. In this case sampling rates of less than 100% are required for practical considerations. It is possible to reduce sampling errors by taking larger samples. However, beyond certain limits this is not worthwhile because the range of uncertainty will only diminish significantly with very large increases in the sample and also because possible errors due to other causes are likely to become more important.

In addition to these sources of error, the accuracy of a trip matrix obtained using an indirect method will be subject to the following additional ones:

(6) **Calibration errors:** Some items of the data required such as road lengths can be measured very accurately, whereas other items such as employment may only be estimates. This type of error occurs when inaccuracies in the data required for the calibration of parameter values of transport models lead to wrong parameter values. It may also be due to the use of an inadequate and inaccurate calibration procedure.

(7) **Mis-specification errors:** Forecasting models are based on fairly simplified representations of human behaviour even if they appear to be complex. They are, therefore, almost bound to be incorrectly, or

inadequately specified. This type of error occurs when the assumed model for trip making behaviour does not conform to the real trip making behavior accurately. Hence, any approach will suffer from this type of error, although not all to the same extent. Errors on the form of equations used, or the omission of important explanatory variables may show up in comparisons with independent validation data but are more likely to emerge over time. The best guide to the adequacy of a model specification is an examination of the residual errors i.e. the difference between observed and modelled values. They should be normally distributed and should not show any bias. It should be possible to explain why the general form of the model makes sense, and why the dependent variable should vary with the explanatory variables.

The only type of error which has a standard theoretical treatment is the error due to the sampling fraction or sample size error. In this study, however, it is not necessary to discuss this issue in detail.

2.4 Simplified methods using traffic counts

As discussed in the previous sections, all of the conventional methods for estimating trip matrices are fairly expensive, involving considerable resources in terms of manpower, time, disruption to trip makers during the survey and lengthy data processing. Moreover, their end products are sometimes short-lived, unreliable and inaccurate.

In the report 'Urban Traffic Models: possibilities for simplification' published by OECD (1974), two main reasons for pursuing simplification in classical conventional types of urban transport modelling were raised. They are: (1) the high cost of collection and analysis of large volumes of data, (2) the extensive effort required to run complex and costly computer models. It was pointed out that the conventional approach is unwieldy, inflexible and slow and it is not certain that the level of detail and accuracy is well matched to the planning phase concerned. The report suggested that a greater research effort is required to develop and improve strategic models requiring less data, and less computer and manpower resources. The report also stressed the need to improve the applicability of the models to relevant

policy questions and pointed out several outstanding problems regarding their internal consistency.

Transport studies using conventional models have been criticized frequently by many reviewers in the literature (Atkins, 1987). Their common criticism includes that conventional methods are inaccurate, inflexible, too complicated, costly, slow, policy-insensitive and lack a sound theoretical basis.

Realizing the dissatisfaction with conventional methods, the method for estimating trip matrices from traffic counts has received great attention in recent years. The estimation method using traffic counts can avoid at least some of the main disadvantages identified in conventional methods. First of all, traffic counts are relatively easy and inexpensive to obtain. They require less manpower and effort to survey, since they do not require preparation of questionnaires or statutory powers. They might be available from other routine work of traffic management and control such as junction design, accident analysis, monitoring of traffic flows, etc. Also, the automatic collection of vehicle counts is now well advanced and traffic counts on links having an automatic counter can be collected without any further survey. They can be processed easily by computer packages. Furthermore, traffic counts are collected without generating any delay or disruptions to vehicles or trip makers.

Secondly, traffic counts are more accurate and reliable than the data used in conventional modelling. In fact, traffic counts are measured at a higher level of aggregation on links and they are less subject to survey errors. Moreover, all vehicles passing through monitoring points are counted and they have smaller sampling errors.

Thirdly, traffic counts do not need lengthy data processing. They are simple to use and the information provided does not need further manipulation.

Fourthly, the estimation method using traffic counts does not go through the four-stage conventional modelling process. It simplifies

four-stage process into a single stage. Thus, the simplified methods are internally consistent. They do not need much of the large volume of data required in four-stage modelling.

Moreover, traffic counts can be used to update an old trip matrix for example, from previous studies or surveys to be consistent with existing flow observations. In other words, the method using traffic counts has a capability to evolve data base continuously. As data sources age, their relevance and reliability become more dubious and publicly unacceptable. Costs of extensive new data-collection exercises would be prohibitive given the limited role of the current applications. It can no longer be argued that such costs are a small proportion of likely future infrastructure spending.

However, despite many advantages of the simplified method using traffic counts, it has some inherent weaknesses which may fail to be recognized by practitioners. The models have no obvious facility to predict future O-D movements, i.e. they cannot take account of future land-use developments or redistribution effects from a proposal. Also, because of the wide range of O-D patterns that can fit a set of traffic counts, a great importance is attached to prior information. Thus if the old transportation study is based on inadequate, poor or biased models, the new estimates will be similarly flawed.

CHAPTER 3. THE ESTIMATION PROBLEM USING TRAFFIC COUNTS

3.1 Introduction

Following the brief introduction of the trip matrix estimation method using traffic counts in the previous chapter, this chapter further describes a mathematical formulation of the estimation problem. The basic idea of the estimation problem is to derive a trip matrix which closely reproduces traffic counts observed on links, when reassigned to the network. Traffic assignment plays an important role in the formulation of this problem, as it relates an estimated matrix to traffic counts.

Section 3.2 reviews various traffic assignment methods available ranging from a simple all or nothing assignment to a more advanced equilibrium assignment. Section 3.3 describes a general mathematical formulation of the matrix estimation problem. Section 3.3 ends by discussing some major difficulties inherent in the estimation problem.

3.2 Review of traffic assignment methods

The traffic assignment problem may be stated as follows:

Consider a road network representing the study area and a trip matrix $\{T_{ij}\}$ representing the number of trips made between origins and destinations in the network. Then, the objective of the traffic assignment is to determine traffic flows on the links of the network by modelling the routes taken by drivers through the network.

Route choice procedures attempt to simulate drivers' behaviour in choosing the routes that drivers' consider best. Different drivers often choose different routes between any given origin and destination, especially in urban areas. There might be numerous factors which affect the route selection of drivers such as:

- time,
- distance,
- monetary cost,
- congestion and/or cost,
- type of road (e.g. motorway vs A-road),
- scenery,
- road works,
- safety,
- sign posts,
- habit.

To include all such behavioural factors in a model is clearly not possible and therefore approximations are necessary. A number of studies have attempted to identify the specific factors which motivate drivers. See, for example, Benshoof (1970), Ratcliffe (1972), Armstrong (1977), Outram and Thompson (1977, 1978), Lumm (1978), and Wootton, Ness and Burton (1981). However this research reviews only two very general reasons why drivers choose different routes:

(1) Difference in individual perceptions of cost known as stochastic effects: Drivers differ in their perception of what constitutes the best routes. For example, some drivers choose the fastest route and some the shortest. Such differences in perception will lead to a division of traffic between two routes.

(2) Capacity restraint or congestion effects: Increased travel costs due to congestion on heavily used links, making some routes less attractive and increasing the number of routes of similar cost between any two nodes.

Both stochastic effects and congestion effects may be modelled into the context of 'travel costs'. We may assume that each trip maker chooses his route so as to minimise his individually perceived travel cost and that trip makers vary in the way in which they perceive these costs. For example, a driver interested in a fastest route equates cost with time, the shortest route equates cost with distance, etc. The modeller must usually attempt to define a travel cost which represents

in some sense an average travel cost taken over all drivers. The most common method is to define cost as a linear combination of time and distance.

Stochastic effects arise because different drivers perceive costs in different ways, whereas capacity restraint arises because costs - and in particular their travel time component - depend on flows. Strictly speaking both effects operate together, particularly in urban areas, and a perfect route choice model would take both into account. However, it appears that stochastic effects are the dominant factor at low levels of traffic flow whilst capacity restraint becomes dominant at higher flows.

Along with the route choice criteria - stochastic effects and congestion effects - traffic assignment methods may be classified into four ways: all or nothing, pure stochastic, Wardrop equilibrium (Wardrop, 1952) and stochastic user equilibrium. For modelling the problem of trip matrix estimation from traffic counts, Robillard (1975) has classified traffic assignment methods into two groups: proportional assignment and non-proportional ones. In this classification, proportional assignment methods satisfy the following conditions:

(1) the total assigned flow on a link is the summation of all the flows assigned if each O-D pair is assigned separately, and

(2) if all the elements of the trip matrix are changed by a certain fraction, then all the assigned flows on each link also changes by the same fraction. For example, if all the entries of the trip matrix are doubled, the assigned flow will double the flow assigned with the original trip matrix.

Any trip assignment process which does not conform with these above two conditions is classified as non-proportional. All or nothing and pure stochastic assignments belong to the proportional assignment group. Wardrop equilibrium assignment and stochastic user equilibrium are non-proportional.

3.2.1 Proportional traffic assignment methods

(1) **All or nothing assignment:** If we assume that all drivers perceive travel costs in an identical fashion and that these costs are fixed independent of flows, then every driver from i to j must choose the same route. In certain circumstances such assumptions may be justified - for example, in a relatively sparse network of uncongested rural roads - but they are unlikely to apply to traffic in urban areas. However, all or nothing assignment is still the simplest and most efficient technique, hence its widespread use. Having found the shortest path - minimum cost - between each origin and destination and loaded the trips onto the network through these, then the total flow for each link can be calculated.

All or nothing assignment is the fastest and simplest method of assignment and useful for simple networks where there are only few alternative O-D paths or little congestion. In addition to this, the use of all-or-nothing assignment provides useful information to the traffic planner in that it represents a 'desire line' assignment which the traffic planner might choose to use to plan new roads. It is also often necessary to use this assignment in considering how schemes, which are designed primarily for peak hour conditions, operate during off-peak. Another important role of all or nothing assignment is to use for other assignment methods. For example, the iterative Frank-Wolfe algorithm for solving the equilibrium assignment problem uses all or nothing assignment to generate a set of auxiliary link flows. This will be reviewed in detail later.

(2) **Pure stochastic assignment:** Pure stochastic assignment retains the assumption of flow-independent costs but takes into account variations in drivers' perceptions on route choice. This type of assignment differs from all or nothing assignment in that it seeks to spread drivers across a range of different routes between each origin and destination, explicitly allowing non-minimum cost routes to be selected. Stochastic assignment is often useful for generating multi-routes in uncongested networks. A number of algorithms have been proposed to do so; see for example, Burrell (1968), Dial (1971), Van Vliet (1976), Florian and Fox

(1976), Daganzo and Sheffi (1977), Van Vliet and Dow (1979) and Sheffi and Powell (1981). Here two most widely used methods of stochastic assignment - **Burrell's** and **Dial's** methods - will be reviewed briefly.

Burrell's assignment method assumes that the travel cost perceived by individual drivers on each link of the network is distributed around the mean cost. The model defines the actual mean link costs together with a form of the distribution about this mean of individually perceived travel costs on each link. The perceived travel costs on each link can then be generated by taking random samples of these link cost distributions. Random numbers, for example based on a rectangular distribution, are used repeatedly to select costs for each link. The model then finds and loads the fastest routes minimising the sum of their perceived travel costs. A number of variants on this basic theme are used in practice, based, for example, on different forms of the distribution. (See, for example, Brooks and Harris, 1972; Mason, 1972).

Burrell assignment satisfies the criterion that cheap routes are used more frequently than expensive ones. It also alleviates the problems associated with parallel routes where an all-or-nothing algorithm tends to load all trips onto one route and none onto parallel routes. However, it does suffer from one major problem in that the flows generated are subject to stochastic fluctuations, so that in comparing two different schemes it is sometimes difficult to distinguish real differences from stochastic ones. The solution is to repeat the random samplings a sufficiently large number of times to reduce the stochastic fluctuations to an acceptable level, but this may lead to unacceptably long computing times.

Dial (1971) proposed a probabilistic multi-route model whereby trips at each node are sub-divided amongst all feasible entry links in a probabilistic manner favouring minimum cost routes. An attractive feature of Dial's assignment is that more costly routes are assigned less traffic. The user of the model has the facility to calibrate it to suit his needs.

3.2.2 Equilibrium traffic assignment methods

Equilibrium, or capacity restrained traffic assignment is based on assumption that the travellers will consider the generalised travel costs including the congestion effects when they choose their routes. Thus, the cost of travelling on a link depends on the level of the flows on all links, not just that link through the speed-flow relationship. However, this study will be interested in special cases that the cost on a link depends only on the flows on that link.

Various methods for capacity restrained traffic assignment have been proposed. Initially, heuristic or approximate solution techniques were developed. Later, several convergent algorithms satisfying Wardrop's equilibrium principle were devised. The most frequently used heuristic methods include repeated all or nothing, incremental loading and iterative loading assignments:

(1) **Repeated all or nothing:** This involves the assignment of the trip matrix according to some pre-defined values of link costs, usually obtained by an all or nothing algorithm. The link costs are then adjusted according to the speed-flow relationship and the trip matrix is reassigned according to the new link costs. This is repeated until the change in link costs becomes sufficiently small. This procedure is not in general convergent.

(2) **Incremental loading:** The idea is to assign the trip matrix onto the network in fractions and after each loading, the costs of the links are adjusted based on the speed-flow relationship. This process is repeated until the whole trip matrix has been assigned. The accuracy of this procedure depends on the number of fractions used.

(3) **Iterative loading:** This method is to assign the full trip matrix onto the network repeatedly and after each trip assignment, resulting link flows are linearly combined with the link flows from the previous iteration. There are different ways of choosing at each iteration but the most popular is due to Smock (1962) who suggested that should be made equal to the reciprocal of the number of iterations. It can be seen

that this results in a small value for after some iterations, ensuring small changes of the flows and costs of links.

The basic principles defining traffic equilibrium conditions on congested networks have been first formally enunciated by Wardrop (1952). His first principle is:

Under equilibrium conditions traffic arranges itself in congested networks so that all routes used between any O-D pair have equal and minimum costs while the cost of any unused route is greater than or equal to this.

Mathematically, Wardrop equilibrium can be expressed as:

$$C_{ijr} = C_{ij}^* \quad \text{if } T_{ijr} > 0 \quad (3.1)$$

$$C_{ijr} > C_{ij}^* \quad \text{if } T_{ijr} = 0 \quad (3.2)$$

where C_{ij}^* is the equilibrium travel cost between zone i and zone j.

Wardrop's first principle is simply a restatement of the basic premise that each driver chooses the route that offers him the minimum perceived cost. That is, traffic distributes itself in such a way that no driver can reduce his travel cost by switching to another route. Thus, under certain circumstances, Wardrop's equilibrium is also referred to as being 'user-optimised' (Dafermos and Sparrow, 1969). However, it was later recognised that Wardrop's equilibrium condition and Dafermos and Sparrow's user-optimised condition do not always identify the same points as equilibrium (Smith, 1984a; Heydecker, 1986). In particular, Smith (1984a) provided an example where Wardrop's equilibrium condition is satisfied but Dafermos and Sparrow's user-optimised condition is not.

Wardrop's second principle states that the distribution of traffic is such that the total travel cost on all routes in the system is minimum. The marginal travel costs on all paths between an O-D pair are equal. Flows which satisfy Wardrop's second principle are often referred as a system optimum.

A major advance in equilibrium assignment was made by Beckmann McGuire and Winsten (1956) who recognised that the Wardrop equilibrium assignment problem is equivalent to a multi-commodity flow problem. They have shown that finding link costs and flows which satisfy Wardrop's first principle is equivalent to finding an optimum solution to the following minimisation problem.

P3.1

$$\text{Min}_{\underline{V}} Z(\underline{V}) = \sum_a \int_0^{V_a} c_a(x) dx \quad (3.3)$$

s.t.

$$\sum_{i,j} \sum_r T_{ijr} \delta_{ijr}^a = V_a \quad (3.4)$$

$$\sum_r T_{ijr} = T_{ij} \quad (3.5)$$

$$T_{ijr} \geq 0 \quad (3.6)$$

where T_{ijr} is the number of trips on route r between i and j , and $\delta_{ijr}^a=1$ if link a is used by route r between i and j , and otherwise, $\delta_{ijr}^a=0$.

The objective function $Z(\underline{V})$ in the problem **P3.1** is a convex function if the link costs $\{c_a(\underline{V})\}$ are separable and non-decreasing functions of the link flows \underline{V} . The problem then becomes a nonlinear programming problem with a convex objective function subject to two sets of linear equality constraints and a set of non-negativity constraints. The simple demonstration that solving the optimization problem yields a Wardrop equilibrium solution can be found in the literature (Van Vliet, 1979; Eash, Janson and Boyce, 1979).

The great advantage of redefining equilibrium assignment as a minimisation problem is that it allows to use a convergent algorithm to find a Wardrop equilibrium solution (Van Vliet, 1979). A number of efficient algorithms suitable for solving even large-scale problems have been proposed and tested. See for example Leventhal, Nemhauser and Trotter (1973), Nguyen (1974b), LeBlanc, Morlok and Pierskalla (1975), Florian and Nguyen (1976), and Van Vliet and Dow (1979). It has been shown that those algorithm are special cases of methods of feasible directions in the mathematical programming (Nguyen, 1974a). Among those,

the Frank-Wolfe algorithm is most commonly used to solve the problem. The Frank-Wolfe algorithm applied in the equilibrium assignment problem may be described as follows (Van Vliet and Dow, 1979).

A3.1

(Iteration 1)

(1) Set all link costs to some predetermined values (generally those corresponding to $V_a=0$).

(2) Build minimum cost trees and assign all T_{ij} to them (all-or-nothing) to produce a set of auxiliary link flows, $F_a^{(0)}$. Set the current main flows $V_a^{(n)} = F_a^{(0)}$ where $n=1$.

(Iteration n)

(3) Alter the link costs in accord with the current main flows, $V_a^{(n)}$, i.e. set:

$$c_a^{(n)} = c_a(V_a^{(n)}) \quad (3.7)$$

(4) Build a minimum cost tree using $c_a^{(n)}$ for each O-D pair and assign all T_{ij} to it by all-or-nothing assignment method. Repeat this for each O-D pair to produce a set of auxiliary link flows, $F_a^{(n)}$.

(5) Generate an improved set of main flows $V_a^{(n+1)}$ as an interior linear combination of old and auxiliary flows:

$$V_a^{(n+1)} = (1-\lambda)V_a^{(n)} + \lambda F_a^{(n)} \quad (3.8)$$

where $0 \leq \lambda \leq 1$, choosing λ so as to minimise the objective function $Z(V^{(n)})$.

(6) Increment n by 1 and return to step (3) unless certain termination conditions have been reached.

The key element of Algorithm A3.1 is the choice of λ . It should be noted that the algorithm does not make direct use of the path flows T_{ijr} and that the choice of λ only refers to combination of link flows and not path flows. One of the important features in equilibrium assignment is that the solutions are unique in terms of path costs, link flows and link costs, but not for path flows and route choice proportions using the individual links. The path flows and route choice proportions can be only extracted heuristically during the assignment process.

Another essential feature of the Frank-Wolfe algorithm is that it is basically iterative, while this algorithm is convergent (Frank and Wolfe, 1956). It may be impossible to find the exact equilibrium flows in a finite amount of computing time (LeBlank et al, 1975). The convergence characteristics of the algorithm are determined by the formulation of the algorithm as well as the initial solution which is used in the application of the algorithm. Rose, Daskin and Keppelman (1988) examined convergence error through an empirical study and reported that the joint selection of an initial solution and a stopping criterion may be important in determining the magnitude of the convergence truncation error.

It is well known that the Frank-Wolfe algorithm gets slower, when it approaches the optimum solution (Guêlat and Marcotte, 1986). A number of modifications have been proposed to speed up convergence (Florian, 1977; LeBlanc et al, 1982; Fukushima, 1984; Arezki and Van Vliet, 1990). Those modifications are basically accomplished by adapting PARTAN (short for parallel tangent) direction in the algorithm (Luenberger, 1986, pp 254-257).

Another different attempt to speed up convergence to the equilibrium solution is to use quantal loading rather than usual all or nothing assignments in the early stages of the equilibrium assignment (Van Vliet and Dow, 1979; Arezki and Van Vliet, 1985). In particular, at the early stages, the use of a quantal loading achieves faster convergence by providing the improved auxiliary flows for the optimum combination.

The Wardrop equilibrium traffic assignment problem may also be formulated in terms of a variational inequality, and this formulation is more general than the Beckmann et al's minimisation formulation (Smith, 1979). Let \underline{V}^* be the vector of equilibrium link flows. Then the variational inequality form of Wardrop's equilibrium states that \underline{V}^* is an equilibrium iff

$$\underline{c}(\underline{V}^*) \cdot (\underline{F} - \underline{V}^*) \geq 0 \quad (3.9)$$

where \underline{F} is any vector of feasible link flows and $\underline{c}(\underline{V}^*)$ is a vector of link costs of the equilibrium flows \underline{V}^* .

Smith (1979) has given conditions which guarantee the existence, uniqueness and stability of Wardrop's equilibrium. Furthermore, Smith (1984b) suggested a descent algorithm for solving a variety of monotone equilibrium traffic assignment problems formulated as a variational inequality.

3.3 The estimation problem

This section describes the general problem of estimating a trip matrix from traffic counts and discusses three fundamental issues - underspecification, errors in traffic counts and congestion effects in traffic assignment - in the estimation problem. More detailed modelling approaches for these issues will be reviewed in Chapter 4.

3.3.1 Description of the problem

We suppose the study area to be represented by a transport network consisting of zones, zone centroids, nodes and links. We suppose the trip making activities between zones for a specified time period to be expressed as the number of trips made per unit time, T_{ij} , whose journeys start at zone i and end at zone j . Furthermore, we introduce the notation of the trip matrix $\underline{T} = \{T_{ij}\}$ in order to express all the trip activities in the network as a single symbol. The rows of the trip matrix correspond to the trips generated within a zone and the columns correspond to the trips attracted to a zone.

The problem of estimating trip matrices in the network can be treated in various ways depending on the input data available and the use which will be made of the trip matrix. In general, the problem can be viewed as containing three important elements: new information such as traffic counts, old information such as out-dated trip matrix and an estimated new trip matrix. These elements are connected by the matrix estimation process. The matrix estimation process is to use traffic counts in order to update an old trip matrix. The problem is then to identify a suitable mechanism to estimate a new trip matrix from various old and new information.

A key issue in the estimation of a trip matrix from traffic counts is how to estimate a trip matrix whose modelled link flows reproduce traffic counts. The most common way to achieve this is to associate traffic counts with assigned link flows from the estimated trip matrix through the process of traffic assignment. Mathematically, this can be expressed as:

$$\underline{V}(\underline{T}) = \underline{\hat{V}} \quad (3.10)$$

where $\underline{V}(\underline{T})$ is the modelled link flows resulting from the assignment of the matrix \underline{T} to the network and $\underline{\hat{V}}$ is a vector of observed link flows.

The way of estimating trip matrices from traffic counts is the inverse process of road traffic assignment. This permits a convenient alternative to the first three stages of the conventional four-stage modelling exercise.

The modelled link flows are generated from traffic assignment, provided that a trip matrix is known. For the time being, we assume that the proportion of trips from origin i to destination j which use a particular link a in the network is known explicitly. We shall use the variable $\underline{P} = \{P_{ij}^a\}$ to express this value.

In general,

$$0 \leq P_{ij}^a \leq 1, a=1,\dots,M, i=1,\dots,N, j=1,\dots,N. \quad (3.11)$$

where the extreme values occur either when the link is not used by any trips from i to j or when it is used by all of them. If we use the proportion P , the equation (3.10) becomes:

$$\underline{PT} = \underline{V} \quad (3.12)$$

or,

$$\sum_{i,j} P_{ij}^a T_{ij} = \underline{V}_a, a=1,\dots,M, i=1,\dots,N, j=1,\dots,N. \quad (3.13)$$

This is M simultaneous linear equations with $N(N-1)$ unknowns. However, in practice, the number of unknown variables is greater than the number of equations, as the number of O-D pairs is normally far greater than the number of links in the network. Thus, the estimation problem is underspecified in most cases and in general there will be more than one solution satisfying the equation, if traffic counts are error-free and mutually consistent.

Consider the simple network depicted in Figure 3.1. This network has two origins (a and b) and two destinations (c and d). The flows on all links are also shown in this figure.

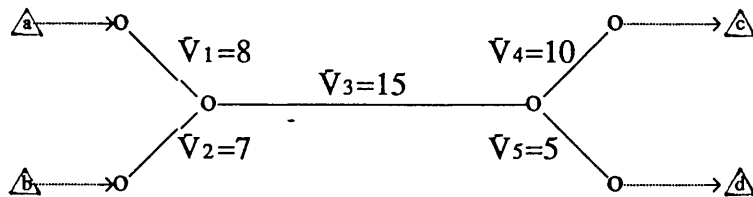


Figure 3.1 Simple network showing an underspecified problem

It is possible to draw a set of linear equations from Figure 3.1 using the path and link flows relationship.

$$T_{ac} + T_{ad} = \bar{V}_1 \quad (3.14)$$

$$T_{bc} + T_{bd} = \bar{V}_2 \quad (3.15)$$

$$T_{ac} + T_{ad} + T_{bc} + T_{bd} = \bar{V}_3 \quad (3.16)$$

$$T_{ac} + T_{bc} = \bar{V}_4 \quad (3.17)$$

$$T_{ad} + T_{bd} = \bar{V}_5 \quad (3.18)$$

However, it can be shown that only three of five equations are independent (see Section 3.3.3). Therefore, the problem becomes underspecified, since the number of unknowns, 4, is greater than the number of equations, 3.

3.3.2 Treatment of underspecification

In Section 3.3.1, it was observed that the estimation problem is underspecified. Here, the discussion is extended to possible ways of determining a single trip matrix out of the infinite feasible solution set. For that purpose, some extra mechanism or principle is needed to reduce the number of unknowns of the estimation problem so that it becomes fully specified.

Approaches for reducing this underspecification problem have been developed by many researchers. One reasonable way to overcome this problem is to restrict the number of possible solutions by making assumptions about trip making behavior. For example, the most widely used assumption is based on the entropy maximization theory. The entropy maximization theory has been used widely to explain trip making behavior (Wilson, 1967). The more details about various modelling approaches on this issue will be covered in Chapter 4.

Another practical way for treating underspecification is to use old information such as out-dated trip matrices. In this case, it can be considered to update the old trip matrix using the new information.

3.3.3 Treatment of inconsistent link flows

In Section 3.3.1, we assumed that traffic counts are independent and consistent, but in reality, they are not. Certain combinations of

traffic counts might make it impossible to estimate a trip matrix to satisfy them. These problems are discussed in terms of dependence and inconsistency of traffic counts (Willumsen, 1981b).

(1) **Dependence:** Some link counts might be expressed as a linear combination of others. Such counts will fail to add any information, i.e.

$$\hat{V}_a = \sum_{i \in I - \{a\}} \gamma_i \hat{V}_i \quad (3.19)$$

where \hat{V}_i is observed link flows in link i , $i \in I$,
 γ_i is constant coefficient of link flows V_i ,
 I is the set of observed links in the network.

(2) **Inconsistency:** Two sources for inconsistencies in traffic counts in the matrix estimation are identified. The first one is that counting errors and asynchronous counting - often traffic counts are obtained on different occasions (hours, days, weeks) - are likely to lead to inconsistency in the link flows. In this case, the set of link flows may fail to keep the principle of the conservation of the link flows, i.e.

$$\sum_l \hat{V}_{lm} \neq \sum_k \hat{V}_{mk} \quad (3.20)$$

where \hat{V}_{lm} is observed link flows from node l ,
 \hat{V}_{mk} is observed link flows to node k .

For example, in Figure 3.1, if traffic counts in the link 3 were given as $\hat{V}_3=20$ instead of $\hat{V}_3=15$, the conservation of the link flows is not met and they become inconsistent.

The second source is a mismatch between the assumed traffic assignment model and traffic counts. For example, an assignment model may allocate no trips on a link having traffic counts. In these conditions, there will be no trip matrix reproducing the traffic counts on the link using that assignment model.

The existence of inconsistencies in traffic counts might lead to there being no feasible solution. There are two possible ways to resolve this difficulty. The first way is to correct errors before estimating trip matrices. Although it is possible to have independent flows, the inconsistency between link flows are found to be more difficult to correct. The second way is to accommodate these errors within the formulation of the estimation problem. Although the first approach always has an advantage to produce a feasible solution, it might be far from true values. The second one tries to produce a solution close to a feasible solution. Further details on this issue will be reviewed in Chapters 4 and 5.

3.3.4 Treatment of traffic congestion effects

As reviewed in Section 3.3.1, traffic assignment plays an important role in the estimation of a trip matrix from traffic counts. The estimated trip matrix can be only constrained by traffic counts through the traffic assignment process. Thus, the use of an appropriate traffic assignment method is important in the determination of an estimated trip matrix.

As discussed in Section 3.2, there are two main types of traffic assignment methods available. The first type, known as proportional assignment, including all or nothing assignment and stochastic assignment, does not consider any congestion effects on the choice of routes. The second case, capacity restrained assignment including the Wardrop equilibrium assignment is based on the assumption that travellers will consider the generalised costs including any congestion effects when they choose their routes.

In section 3.3.1, we introduced variable of the assignment proportions \underline{P} with an assumption of being known explicitly. In this section, we will have a more general discussion on the use of this variable.

In the case of proportional assignment, the level of trip demand has no effects on the assignment proportions \underline{P} . It is possible to

identify the proportions \underline{P} independently of the trip matrix estimation process. This approach is efficient and easy to use. However, there is a considerable amount of empirical evidence that this approach cannot explain the route choices of all drivers in urban networks. For example, see Van Vliet (1976), Outram and Thompson (1978), and Van Vliet and Dow (1979).

In the case of capacity restrained equilibrium assignment, the assignment proportions of the trips choosing each route in the network are not constant when the level of travel demand varies. The partial derivative of \underline{P} with respect to \underline{T} is not in general equal to zero. Thus,

$$\frac{\partial \underline{P}}{\partial \underline{T}} \neq \underline{0} \quad (3.21)$$

For example, consider a simple network depicted in Figure 3.2. The network has one origin o , one destination d and two links 1 and 2.

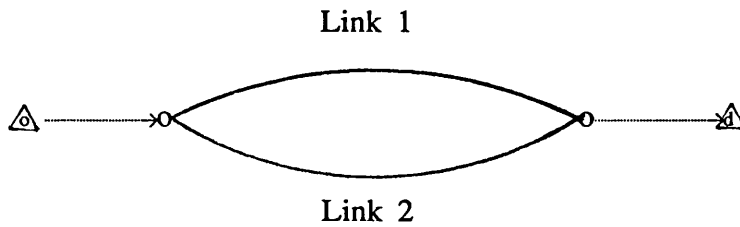


Figure 3.2 Simple network showing variable assignment proportions

We shall assume that the link costs on links 1 and 2 are given as

$$c_1(V_1) = V_1, \quad (3.22)$$

$$c_2(V_2) = K + V_2, \text{ where } K \text{ is a constant.} \quad (3.23)$$

Then, when the equilibrium conditions are met, we can obtain the following assignment proportions of the trips using each of links 1 and 2. When the demand T_{od} is less than K , all the trips use Link 1. As the

demand T_{od} becomes greater than or equal to K , the trips are split into Link 1 and Link 2.

$$\text{If } 0 \leq T_{od} \leq K, \quad P_{od}^1 = 1 \quad \text{and} \quad P_{od}^2 = 0 \quad (3.24)$$

$$\text{If } K \leq T_{od}, \quad P_{od}^1 = \frac{T_{od} + K}{2T_{od}} \quad \text{and} \quad P_{od}^2 = \frac{T_{od} - K}{2T_{od}} \quad (3.25)$$

These results show that the assignment proportions \underline{P} in the capacity restrained equilibrium assignment are not constant as the level of travel demand varies.

Also, the assignment proportion \underline{P} are not always differentiable. It is not possible to identify the proportions \underline{P} independently of the matrix estimation process: as a trip matrix changes, the proportions \underline{P} also change. Furthermore, the proportions \underline{P} are not uniquely determined by the equilibrium assignment process. For these reasons, the estimation problem combined with capacity restrained assignment becomes more difficult to solve.

In Section 3.3.2, it was noted that the estimation problem is underspecified even with a fixed set of the assignment proportion. Now, the problem is further underspecified, since the assignment proportions \underline{P} are also unknown in addition to the $N(N-1)$ unknowns of the trip matrix.

CHAPTER 4. REVIEW OF RELATED METHODS

4.1 Introduction

As mentioned in Chapter 2, conventional methods for estimating trip matrices are economically expensive due to a difficulty of obtaining the required input data and more importantly their end products are likely to be short-lived and unreliable. As an alternative to conventional methods, simplified methods using traffic counts have received great attractions because of their practical advantages including an inexpensive acquisition of traffic counts. Since the potential advantages of simplified methods using traffic counts have been recognized, various modelling approaches have been proposed. Some of them have been already shown to be useful.

The main objective of this chapter is to provide a detailed and up-to-date review of the various methods for estimating trip matrices from traffic counts. Particular attention will be given to the methods which work with equilibrium traffic assignment.

First of all, Section 4.2 classifies various existing methods for estimating trip matrices from traffic counts according to the type of traffic assignment methods and the way of tackling the underspecification problem. This helps to carry out a systematic review. Following the classification, Section 4.3 reviews methods which use proportional traffic assignment methods. Section 4.3 is further divided into three sub-sections devoted to: methods for calibrating synthetic demand models, methods based on information theory, and methods based on statistical inference. Section 4.4 reviews methods which use equilibrium traffic assignment methods. Again, it is further divided into three sub-sections, each for a particular method: Willumsen's method, Nguyen's method and Fisk and Boyce's method.

4.2 Classification of existing estimation methods

Based on a general description of the estimation problem described in Section 3.3.1, this section classifies various methods for estimating trip matrices from traffic counts before a further detailed review. At the least, this usefully initiates the discussion of some common issues and ideas to proceed in a more or less systematic way.

For the first time, Willumsen (1978a; 1981b) provided a review for methods to estimate trip matrices from traffic counts. Willumsen identified the different behavioral assumptions used to tackle the underspecification of the estimation problem and then these have been used to group together similar methods. The estimation methods have been classified into three groups. A first group encompasses models which require the assumption that some form of the gravity model is capable of explaining most of the trip making behaviour in the study area. A second group exploits certain properties of equilibrium assignment techniques to provide an estimate of the trip matrix. Finally, a third group uses entropy maximizing or related techniques to provide the most likely estimate of the trip matrix. However, he argued that this classification, as any other, is not perfect and models exist with characteristics of more than one of these groups. For example, models in groups 1 and 3 assume proportional assignment to be sufficiently realistic but some of them can be extended to include, at least partially, capacity restraint effects.

Later, another good review for the modelling and the algorithmic development for the estimation methods was carried out by Nguyen (1984). The basic idea used to classify the estimation methods in his paper was similar to Willumsen's one. First of all, he introduced two basic formulations for estimating trip matrices from traffic counts. The first one was associated with the calibration of demand models and the other with the matrix estimation models mainly based on the maximum entropy formalism. Next, these models have been further divided into three cases: single route networks, multi-route networks and equilibrium-based networks. The cases of single route networks and multi-route networks are using the proportional assignment methods for deriving a fixed set

of assignment proportions \underline{P} . The case of equilibrium-based networks are using the equilibrium traffic assignment. In this case, the formulation is to integrate trip matrix estimation and equilibrium traffic assignment without using assignment proportions \underline{P} as these are unknown. Again, he further divided the equilibrium-based estimation methods into two cases: using complete set of traffic counts and incomplete set of traffic counts. By comparison with Willumsen's review, Nguyen placed more efforts on the role of traffic assignment taking into account the new development of the methods using equilibrium traffic assignment methods.

Recently, Cascetta and Nguyen (1988) provided a general review of methods which apply techniques of statistical inference to estimate of trip matrices from traffic counts. Using a generic traffic assignment map, various statistical approaches have been classified into three basic cases: maximum-likelihood estimation method, generalized least square estimation method, and Bayesian approach. Cascetta's review was to reflect a recent development of statistical inference techniques for estimating trip matrices from traffic counts which has occurred, since the previous reviews by Willumsen and Nguyen.

Taking into account these three previous reviews and the main objectives of this research, estimation methods are here classified in the following order. First, the estimation methods are classified according to the type of traffic assignment methods - methods using proportional assignment and methods using equilibrium traffic assignment. Then, the methods using proportional assignment are further classified according to the methods used to accommodate the underspecification - methods for calibrating synthetic demand models, methods based on information theory and methods including statistical inference. The methods using equilibrium assignment are further divided into: Willumsen's method, Nguyen's method and Fisk and Boyce's method.

The main objective of this research is to develop a new method for estimating trip matrices from traffic counts, when congestion effects in networks are considered to be important. Therefore, the methods using equilibrium traffic assignment will be reviewed in greater detail than

the methods using the proportional assignment.

4.3 Methods using proportional traffic assignment

As discussed in Section 3.3, it is possible to determine a set of fixed assignment proportions \underline{P} under the assumption of using proportional traffic assignment. The combination of observed link flows and the assignment proportions \underline{P} leads to a formulation of simultaneous linear equations. The observed flows \underline{V}_a on link a are equal to be the summation of the contributions of all trips between zones to that link. For convenience, Equation (3.13) is repeated again.

$$\sum_{i,j} P_{ij}^a T_{ij} = \underline{V}_a, \quad a=1,\dots,M, \quad i=1,\dots,N, \quad j=1,\dots,N. \quad (3.13)$$

It has already been noted that Equation (3.13) is normally underspecified, as the number of the unknowns $N(N-1)$ is far greater than the number of equations M . Because of this, Equation (3.13) requires a further mechanism in order to determine a single trip matrix from the many feasible ones. Nguyen (1984) stated the following two general formulations for tackling this underspecification. When the demand model is calibrated from traffic counts, the resulting formulation is:

P4.1

$$\text{Min}_{\underline{\beta}} F(\underline{V}(\underline{\beta}), \underline{\hat{V}}) \quad (4.1)$$

s.t.

$$\underline{PT} = \underline{V}(\underline{\beta}) \quad (4.2)$$

$$T_{ij} = D(O_i, D_j, \underline{\beta}, C_{ij}) \quad (4.3)$$

$$\sum_j T_{ij} = O_i \quad i=1,\dots,N \quad (4.4)$$

$$\sum_i T_{ij} = D_j \quad j=1,\dots,N \quad (4.5)$$

$$T_{ij} \geq 0 \quad (4.6)$$

where $F(\underline{V}(\underline{\beta}), \underline{\hat{V}})$ is an objective function, $V_a(\underline{\beta})$ is the modelled flow on link a , $\underline{\beta}$ is a vector of parameters of the demand model, and C_{ij} is a measure of the separation between zone i and zone j .

When the trip matrix \underline{T} is estimated directly from traffic counts, the resulting formulation is:

P4.2

$$\text{Min}_{\underline{T}} F(\underline{T}, \underline{t}) \quad (4.7)$$

s.t.

$$\underline{PT} = \underline{V} \quad (4.8)$$

$$T_{ij} \geq 0 \quad (4.9)$$

where $F(\underline{T}, \underline{t})$ is an objective function and \underline{t} is the prior trip matrix.

The objective functions used in Problem **P4.1** and **P4.2** may be any suitable distance metric, either between the modelled and observed link flows for Problem **P4.1**, or between the estimated \underline{T} and the prior trip matrix \underline{t} for Problem **P4.2**. For Problem **P4.1**, the most frequently used objective function is a least squares function and for Problem **P4.2**, the most frequently used ones are a least squares function or an entropy function.

4.3.1 Methods for calibrating demand models

Methods for calibrating demand models from traffic counts can be further classified, according to the type of the models they use, into two sub-groups: gravity model based and direct demand model based. In the following, each of these is reviewed and finally a comment is given.

4.3.1.1 Methods based on the gravity model

The gravity model was one of the first mathematical models used for making trip making behaviour in a study area. Because of its simplicity and a certain intuitive appeal, it has received great attention from social scientists and engineers. It is not surprising then that the first ideas on estimating a trip matrix from traffic counts were based on the gravity model.

The gravity model assumes that the trip making behaviour in the area of interest can be explained in terms of three types of factors: trip generation or origin factors, trip attraction or destination factors and separation or travel cost factors. That is, the number of trips from each origin to each destination is directly proportional to traffic generating and attracting factors of the zones and inversely proportional to the separation or travel cost factors between the zones. A common form of the gravity model (Wilson, 1967) is then

$$T_{ij} = A_i O_i B_j D_j f(C_{ij}) \quad (4.10)$$

where A_i and B_j are balancing factors, O_i and D_j are the trip ends, and the function of $f(C_{ij})$ and generalized costs C_{ij} represent the separation or deterrence factor between the zones.

The gravity model can be specified with different degree of sophistication. For example, the different form of the functions $f(C_{ij})$ can be used to model the deterrence. Once a gravity model has been specified, the parameters of the model can be calibrated so that a measure of the errors between the modelled link flows and observed ones is minimized.

The first attempt of this kind based on traffic counts was put forward by Low (1972). He specified a gravity type functional form from which trip probability factors between the zones are calculated:

$$X_{ij}^{(m)} = O_i^{(m)} D_j^{(m)} C_{ij}^{-k} \quad (4.11)$$

where

- $X_{ij}^{(m)}$ is the interzonal trip probability factor between zone i and j for purpose m ,
- $O_i^{(m)}$ and $D_j^{(m)}$ are the trip end totals for purpose m ,
- C_{ij}^{-k} is the deterrence function.

The modelled link flows were generated from the trip probability factors through all or nothing assignment:

$$X_m^a = \sum_{i,j} P_{ij}^a X_{ij}^{(m)} \quad a=1,\dots,M, i=1,\dots,N, j=1,\dots,N. \quad (4.12)$$

where

X_m^a is the modelled link flows for purpose m in link a ,
 $X_{ij}^{(m)}$ is replaced by Equation (4.11).

Finally, a linear model for each observed link has been proposed and calibrated from traffic counts by multiple regression analysis techniques. The proposed linear model was:

$$V_a = b_0 + b_1 X_1^a + b_2 X_2^a + b_3 X_3^a + \dots + b_m X_m^a, \quad a=1,\dots,M. \quad (4.13)$$

where

V_a is the modelled flows in link a for all purposes,
 b_m are the constants to be calibrated.

Equation (4.13) is used to forecast the future link flows by re-estimating trip probability factors from the future trip end data. The estimation procedure has been applied in Monongalia County Transportation Study in West Virginia in 1970-71. The result provided grounds for optimism. Later, Smith and McFarlane (1978) applied Low's approach to the county of Fond du Lac in Central Wisconsin for the purpose of further evaluating Low's approach. The result has showed that the level of accuracy of the model in reproducing base year link flows was certainly within the limits achieved by conventional methods.

The estimation procedure proposed by Low effectively combines the four-stage conventional modelling process into a single one. It has been noted that the primary advantage seems to lie in the fact that the approach can be used to produce traffic volume forecasts when the resources required to use conventional methods are lacking.

Other methods similar to Low's have been developed and tested by different researchers. In this thesis, it is not necessary to introduce all of them in detail. Only the main characteristics of the methods which are interesting and relevant to this research will be described.

Other similar estimation methods based on the gravity model have been proposed by Overgaard (OECD, 1974), Robillard (1975), Högberg (1976), and Holm et al (1976). Overgaard's method includes a variable of car ownership levels that acts as a measure for trip making propensity which is lacking in Low's approach. Robillard proposed a simple double factor functional form leading to the least squares problem. The proposed problem is solved by non-linear regression and it does not require a priori data about the generation and attraction power of each zone. The method only uses traffic counts to calibrate the model. The model put forward by Högberg includes a more flexible deterrence function but requires non-linear regression for its calibration. An interesting feature found in Holm et al's method is the use of Smock's (1962) assignment procedure to refine the traffic assignment proportions P iteratively. This represents a first attempt to take into account the dependency of the traffic assignment proportions and the interzonal travelling costs C_{ij} over the trip demand in the estimation problem. The sequential approach similar to Smock's one will be described in more detail together with other methods which use equilibrium traffic assignment in Section 4.4.

Recently, Wills (1986) developed a flexible gravity-opportunities model for trip distribution in which standard forms of the gravity and intervening opportunity models are obtained as special cases of a general gravity-opportunity model. The idea of the intervening-opportunity model is to represent the effects of varying density of trip end opportunities on trip making behaviours, whereas the gravity model is not sufficient in modelling them. Wills tested his models empirically to estimate a trip matrix from traffic counts and showed that a significant improvement is obtained over the gravity model. Following Wills's approach, Tamin (1988) also explored more advanced flexible demand models - gravity, opportunity and gravity-opportunity models - and proposed three different solution methods - non-linear least squares, weighted non-linear least squares, and maximum likelihood methods - to calibrate the demand models. He performed a wide range of the empirical tests and comparisons over the performance of the demand models and solution methods using the data surveyed in the town of Ripon in England. All methods were found to perform satisfactorily since each

calibrated model reproduced the observed trip matrix closely. In particular, it was found that the gravity model and the gravity-opportunity model with the non-linear least squares method produce the best fit with the observed data in Ripon.

4.3.1.2 Methods based on direct demand models

Unlike the gravity model, in which the zonal trip end data are prepared separately, in direct demand models all the elements of generation, attraction, distribution between the zones are combined into a single model (Domencich and McFadden, 1975). In general, direct demand models are based on the assumption that the aggregate volume of trips between the zones may be directly modelled as a functional form of the following variables: socio-economic factors such as population and employment, accessibility factors such as centrality and deterrence factors between the zones. Early examples include Domencich et al (1968) and Quandt and Baumol (1966). For instance, two of the more common forms (Carey, Hendrickson and Siddharthan, 1981) are a linear form:

$$T_{ij} = b_0 + b_1 X_i + b_2 X_j + b_3 C_{ij} \quad (4.14)$$

where X_i and X_j are socio-economic characteristics of the two zones and b_i ($0 \leq i \leq 3$) are the parameters to be calibrated,

and a multiplicative form:

$$T_{ij} = b_0 X_i^{b_1} X_j^{b_2} C_{ij}^{b_3} \quad (4.15)$$

As with the methods based on the gravity model, the parameters of direct demand models are calibrated by minimising the sum of squared errors between the modelled and observed flows. That is:

P4.3

$$\text{Min}_b (\bar{V} - \underline{PT})'(\bar{V} - \underline{PT}) \quad (4.16)$$

where \underline{T} is replaced by Equations (4.14) or (4.15).

Wills (1978) proposed linear and nonlinear demand models and then calibrated them by the method described above. Carey et al (1981) presented a constrained least-square estimation model using a linear direct demand model and an incomplete data set. It is interesting to note that the model also includes the observed O-D flows in addition to the observed link flows. Later, Carey and Revelli (1986) extended the model by adding statistical inference properties. The work was also extended using a non-linear direct demand model and including inequality constraints. The formulation is more flexible, as any available information may be included in the form of constraints on the optimal process.

4.3.1.3 Discussion

The methods based on the gravity model or the direct demand model certainly have an intuitive appeal and most of them are fairly simple to apply.

In general, the number of parameters calibrated is far less than the number of traffic counts, so the demand models are overspecified. In order to overcome this difficulty, they are fitted over traffic counts by the least squares method. This leads to a feature that a trip matrix estimated is heavily structured over the values of the few parameters. Methods based on demand models are unlikely to be successful when applied in urban areas, where travel patterns tend to be less structured. The gravity model or the direct demand model have more sound grounds for applications to inter-city transport rather than urban transport. Unlike urban travel patterns, intercity travel patterns are more likely to be characterised by several influential factors such as population, level of economic activity, land use, travel deterrence between the zones, etc.

Except for Robillard's approach, most of the methods reviewed in this section require the trip end data for the gravity model or the attraction and generation capacities for the direct demand models to be prepared along with the models selected before calibration. This contrasts with the methods reviewed in the following two sections, which

require only traffic counts.

As earlier mentioned in Low's approach, one of the main advantages of this type of approach is that it only needs simple and low-cost data, whereas conventional methods require large amounts of data and considerable technical resources. For instance, this type of approach might be especially useful in developing countries where there are many difficulties for the use of conventional methods such as the fast rate of change, poor quality of data and data collection practice, and lack of technical experience.

4.3.2 Methods based on information theory

In this type of estimation method, underspecified problems are solved by enforcing extra principles based on information theory and the most likely trip matrix is estimated from feasible ones. From a review of various literatures, the most frequently used principles for this purpose are: maximum entropy and minimum information. We still assume that traffic assignment proportions \underline{P} are independent from the traffic demand to be estimated and can be determined exogenously. Related methods using these two principles will be reviewed in this section.

4.3.2.1 Methods using the maximum entropy principle

The application of the maximum entropy principle to transport and regional planning problems was initiated by Wilson (1967; 1970). Its best known application is in the derivation of a doubly-constrained gravity model by maximizing a measure of entropy subject to trip end and total cost constraints. The derivation of the entropy functions will be given in detail later in Chapter 5. The problem formulated by Wilson (1970) was:

P4.4a

$$\text{Max}_{\underline{T}} S(\underline{T}) = - \sum_{i, j} T_{ij} (\log_e T_{ij} - 1) \quad (4.17a)$$

s.t.

$$\sum_j T_{ij} = O_i, \quad i=1, \dots, N \quad (4.17b)$$

$$\sum_i T_{ij} = D_j, \quad j=1, \dots, N \quad (4.17c)$$

$$\sum_{i, j} C_{ij} T_{ij} = C \quad (4.17d)$$

$$T_{ij} \geq 0 \quad (4.17e)$$

where C is the total (unknown) travel cost in the network.

Solving Problem P4.4a by forming the Lagrangian results in the doubly-constrained gravity model with the exponential deterrence function $\exp(-\beta C_{ij})$. The general form of the gravity model has already given in Equations (2.2) and (4.10) in Sections 2.3.2 and 4.3.1.

$$T_{ij} = A_i O_i B_j D_j \exp(-\beta C_{ij}) \quad (4.17f)$$

where A_i and B_j are balancing factors related to the trip ends constraints (4.17b) and (4.17c) and calculated as

$$A_i = \frac{1}{\sum_j B_j D_j \exp(-\beta C_{ij})} \quad (4.17g)$$

$$B_j = \frac{1}{\sum_i A_i O_i \exp(-\beta C_{ij})} \quad (4.17h)$$

and β is the Lagrange multiplier related to the total cost constraint (4.17d) and represents the degree of the perception to the generalised travel cost, C_{ij} .

Willumsen (1978a; 1978b; 1981a) developed a model based on the maximum entropy principle to estimate a trip matrix from traffic counts. Willumsen's model is known as the Maximum Entropy Matrix Estimation (ME2) model. The ME2 model estimates a trip matrix consistent with traffic counts and any prior information available. The derivation of the ME2 model parallels the derivation of the gravity model but replacing trip ends and total cost constraints with constraints

associated with traffic counts. The problem then formulated only with traffic counts is

P4.4

$$\text{Max}_{\underline{T}} S_1(\underline{T}) = - \sum_{i,j} T_{ij} (\log_e T_{ij} - 1) \quad (4.17)$$

s.t.

$$\sum_{i,j} P_{ij}^a T_{ij} = \bar{V}_a, \quad a=1,\dots,M, i=1,\dots,N, j=1,\dots,N. \quad (4.18)$$

$$T_{ij} \geq 0 \quad (4.19)$$

Problem P4.4 is one of convex programming since it consists of a strictly concave and continuous function subject to linear constraints. Any primal convex programming techniques may be used. The formal solution is obtained (Willumsen, 1981a) by forming the Lagrangian

$$L(\underline{T}, \underline{\lambda}) = - \sum_{i,j} T_{ij} (\log_e T_{ij} - 1) - \sum_{a=1}^M \lambda_a (\sum_{i,j} P_{ij}^a T_{ij} - \bar{V}_a) \quad (4.20)$$

where λ_a is the Lagrange multiplier associated with link count a .

Differentiating this with respect to T_{ij} gives:

$$\frac{\partial L(\underline{T}, \underline{\lambda})}{\partial T_{ij}} = -\log_e T_{ij} - \sum_{a=1}^M \lambda_a P_{ij}^a \quad (4.21)$$

Thus, for stationarity,

$$T_{ij} = \exp(-\sum_{a=1}^M \lambda_a P_{ij}^a) \quad (4.22)$$

and by making

$$\exp(-\lambda_a) = X_a \quad (4.23)$$

we finally obtain

$$T_{ij} = \prod_{a=1}^M X_a^{P_{ij}^a} \quad (4.24)$$

In order to determine the optimum solution \underline{T} from Equation (4.24), it is necessary first to determine values for the Lagrange multipliers $\underline{\lambda}$. This can be done by substituting Equation (4.24) into the link flow constraints (4.18) and solving M non-linear equations simultaneously for the M Lagrange multipliers. In practice though, this approach requires considerable computer memory for practical size problems since the whole of the array for assignment proportions \underline{P} must be available at all times. This leads to the consideration of efficient iterative row generation techniques in which only one constraint is considered at a time. This method known as the multi-proportional procedure has been studied extensively by Murchland (1977; 1978). Later, Lamond and Stewart (1981) showed that Kruthof's double factor method (1937), Evans and Kirby's tri-proportional method (1974) and Murchland's multi-proportional method (1977) are special cases of efficient row generation balancing method studied by Bregman (1967). Bregman proved that the algorithm converges to a unique solution provided that the constraints are mutually consistent. The algorithm adopted by Willumsen (1981a) is as follows. The following algorithm requires setting up a list of observed links from $a=1$ to M .

A4.1

(step 1) Obtain, using a suitable assignment method, the values of assignment proportions \underline{P} and set the number of iterations $n=0$.

(step 2) Set $X_a^{(n)}=1$ for all links.

(step 3) Set the counter $a=0$.

(step 4) Increase the link counter a by one. Take link a and calculate modelled flows.

$$V_a^{(n)} = \sum_{i,j} P_{ij}^a \left(\prod_{k=1}^M X_k^{(n) P_{ij}^k} \right) \quad (4.25)$$

Replace $X_a^{(n)}$ by $X_a^{(n+1)} = X_a^{(n)} Y_a$ for each link a where Y_a is obtained by solving

$$\bar{V}_a = \sum_{i,j} P_{ij}^a \left(\prod_{k=1}^M X_k^{(n) P_{ij}^k} \right) Y_a^{P_{ij}^a} \quad (4.26)$$

Equation (4.26) may be solved by a unidimensional Newton-Raphson method.

(step 5) If \bar{a} is less than L , proceed to (step 4).

Otherwise, move to (step 6).

(step 6) If the difference between $\{V_a^{(n)}\}$ and $\{\bar{V}_a\}$ is within the convergence limit, calculate $\{T_{ij}\}$ using Equation (4.24) and terminate. Otherwise, set $n=n+1$ and return to (step 3).

The ME2 model can also be extended to make use of prior information such as out-dated trip matrices. The resulting problem is:

P4.5

$$\text{Max}_{\underline{T}} S_1(\underline{T}, \underline{t}) = - \sum_{i,j} T_{ij} (\log_e (T_{ij}/t_{ij}) - 1) \quad (4.27)$$

s.t.

$$\sum_{i,j} P_{ij}^a T_{ij} = \bar{V}_a, \quad a=1,\dots,M, i=1,\dots,N, j=1,\dots,N. \quad (4.28)$$

$$T_{ij} \geq 0 \quad (4.29)$$

When no prior information is available, one could plausibly set $t_{ij}=1$ for all i and j , Problem P4.5 reverts to Problem P4.4. Again, Problem P4.5 is a convex optimization problem with linear constraints. It can be solved by the same process described earlier. In this case, instead of (4.24), we obtain for Problem P4.5:

$$T_{ij} = t_{ij} \prod_{a=1}^M X_a^{P_{ij}^a} \quad (4.30)$$

where $X_a = \exp(-\lambda_a)$.

Equations (4.28) and (4.30) can be solved by the same multi-proportion procedure described above.

From Equations (4.24) and (4.30), it is interesting to note that the factor X_a is associated with the contribution of the observed link flows on link a to the formation of the trip matrix in Problem P4.4 or the modification of the trip matrix in Problem P4.5. The factor X_a plays a role analogous to the balancing factors in a doubly-constrained gravity model. In both cases, this contribution is weighted by the exponent P_{ij}^a representing the proportion of trips between each O-D pair that use link a .

As described earlier in Section 3.3.3, in practice traffic counts are unlikely to be error-free. Errors in link flows occur partly because of counting errors and partly because counts may be carried out at different times or on different days. Two sources of errors in traffic counts have been identified. The first one, linearly dependent link flows, can be easily detected and removed. The second one, inconsistency in link flows, prevent the estimation problems from being feasible. There are two general approaches for dealing with inconsistent link flows. The first one is to develop trip matrix estimation methods in which inconsistent link flows are accommodated. This type of the estimation method will be reviewed in the next section on methods using statistical inference techniques. The second approach is to remove inconsistency and generate a better estimation of the link flows by using the maximum likelihood method before trip matrix estimation, thus adding the extra information. This approach has been studied by Van Zuylen and Willumsen (1980). Later, Van Zuylen and Branston (1982) extended this approach to the case when more than one count is available on some links of the network.

The ME2 model and the associated solution method have been tested empirically using the data collected by TRRL (Leonard and Tough, 1979) from a comprehensive vehicle license plate survey in the central area of Reading in England (Willumsen, 1981a; Van Vliet and Willumsen, 1981). The survey period covered four successive afternoon periods (16:10 to 18:10) starting on Monday 18 October 1976. The road system in the study area was coded into a network with 39 trip end zones, 80 nodes and 159 one-way links. The observations were then processed through the network to produce:

- an observed or sampled trip matrix,
- a set of observed paths and route choice proportions, and
- a set of 159 observed link volumes

in terms of passenger car units (pcu's) for each of the four periods surveyed. These samples were not grossed up and the tests referred to this unexpanded level which represented a 7 per cent sample of all the movements in the area. However, in order to maintain a correct treatment of delays wherever speed-flow relationships were used, the link flows were expanded to represent 100 per cent volumes. In particular, instead of modelled traffic assignment proportions P , the observed route choice proportions could be used to carry out the sensitivity tests of the ME2 model to the effects of the assignment methods.

Because of a close relationship between the ME2 model and this research, the detailed results of the validation tests of the ME2 model are here described. The following major conclusions from the results have been made (Van Vliet and Willumsen, 1981).

(1) Comparison of observed trip matrices and link flows on different days as shown in Table 4.1: The day-to-day variations in the samples of observed trip matrices obtained at a high level of detail are relatively high, implying that a true trip matrix is extremely difficult to measure. It is noted that the variations are much higher at the trip matrix level than at the link flow level as some of the variations average out at the more aggregate level.

Table 4.1 Comparison of observed trip matrices and link flows on different days (taken from Van Vliet and Willumsen (1981))

Indicator	Monday - Tuesday		Tuesday - Wednesday	
	Matrix	Flow	Matrix	Flow
RMSE*	1.9	19.9	1.7	19.1
%RMSE*	160.	21.	135.	20.

* Root Mean Square Error of two sets of quantities T_{ij} and \bar{T}_{ij} is given by $RMSE = \left[\frac{1}{N(N-1)} \sum_{i,j} (T_{ij} - \bar{T}_{ij})^2 \right]^{1/2}$, and %RMSE is expressed as a

$$\text{proportion of the mean, \%RMSE} = \frac{\text{RMSE}}{\text{T}_{\text{mean}}} \times 100(\%), \text{ where}$$

$$\text{T}_{\text{mean}} = \frac{1}{N} \sum_{i,j} T_{ij}.$$

(2) Comparison of observed vs estimated trip matrices using known route choice proportions as shown in Table 4.2: Trip matrices estimated by ME2 using observed routes - so that one important source of error is effectively eliminated - are not very close to the observed trip matrices, indicating that traffic counts alone are not sufficient to estimate trip matrices. In particular, the use of trip end information appears to be valuable to improve the estimated matrix.

Table 4.2 Comparison of observed and estimated trip matrices using known route choice and all or nothing assignment on Tuesday (taken from Van Vliet and Willumsen (1981))

Assignment Indicator	Known route choice		All or nothing	
	no trip ends	with trip ends	no trip ends	with trip ends
RMSE	1.8	1.7	2.1	2.1
%RMSE	149	134	172	169

(3) Comparison of observed and estimated trip matrices using all or nothing assignment as shown in Table 4.2: even with a simple assignment model, all-or-nothing which we know to be fraught with errors, ME2 still yields matrices which are only marginally worse than the observed day-to-day variations.

(4) Finally, it was concluded that ME2 appears to be a reasonably robust model, as the estimated trip matrices are roughly speaking within the day-to-day variations of the observed trip matrices indicating that the model does indeed give reasonable answers.

Matzoros, Van Vliet, Randle and Weston (1987) reported the results from another interesting validation test on the ME2 method carried out using the data collected by Greater Manchester Council before and after

the introduction of a pedestrian scheme implemented in the center of Manchester in November 1980. The main objective of this test was not to evaluate ME2 directly but rather to validate the use of the micro-simulation model SATURN (Van Vliet, 1987) and the ME2 model in the evaluation of traffic management measures. A network and trip matrix for the 'before' network were set up using only data available before introduction of the pedestrian scheme. This was then used to estimate the impacts of that scheme and these estimates were compared with the actual outcome. The network and trip matrix were supplied by GMC. The trip matrix for the weekday morning peak hour (08:00 to 09:00) was obtained from a number of separate sources and collected at considerable cost. The network taken from GMC was recoded to SATURN requirements and recalibrated. Calibration and evaluation were carried out primarily on the ability of SATURN to reproduce observed traffic counts using 73 'before' counts and 35 'after' counts. A new trip matrix was estimated using the 'before' network and 73 'before' traffic counts as data for the ME2 model. This new trip matrix was assigned to the 'before' and 'after' networks to estimate the link flows. These modelled link flows were then compared with the observed ones.

Table 4.3 Comparison of the observed and modelled flows given by the mean absolute difference divided by the average observed flow expressed as a percentage (taken from Matzoros et al (1987))

Network used	Trip Matrix	
	As surveyed by GMC	As estimated by ME2
Before Network	27.0	8.5
After Network	26.2	13.0

Matzoros et al found that as shown in Table 4.3, the use of the SATURN and ME2 models approximately halves the errors in estimating the link flows in the 'before' and 'after' networks by comparison with the ones estimated from the conventionally surveyed trip matrix. In

particular, they ascribed most of the credit for the improvement in fit in terms of flows to the ME2 model. However, the ME2 method is designed to reproduce the observed link flows used in the estimation exactly. This means that the reduction in errors in link flows does not in itself prove anything about the greater validity of the ME2 model over conventional ones. Also, in cases where the traffic management measures taken have little impacts on traffic flows, the good fit between the observed and modelled link flows in the 'after' network does not indicate the validity of the ME2 model, either. Furthermore, the links used for comparison and validation should be independent of those used for matrix estimation so that any prejudice for the ME2 model may not be given.

In spite of the reasonable performance of the ME2 model, it still has some inherent weaknesses which require further investigations. One of the major weaknesses in the ME2 model is that it assumes fixed route choice proportions P which are invariant with traffic demand. This is far from the reality in congested networks. This will be reviewed in greater detail in Section 4.4 along with other methods which use equilibrium traffic assignment.

Another weakness of the ME2 model is that the matrices estimated by it are strongly dependent upon the accuracy of the prior trip matrices (Robertson, 1984; Atkins, 1987). This is partly because traffic counts alone are not sufficient to update a prior trip matrix. This difficulty may be more or less removed in either of two ways. The first one is to use of better prior estimate of the trip matrix t either from an out-dated trip matrix, a larger study or even a travel demand model such as a simple gravity model. The second one is to use other additional information in the form of extra linear constraints such as trip end data or a trip length distribution to improve the accuracy of the estimated trip matrices (Willumsen, 1981a; Van Vliet and Willumsen, 1981).

Robertson (1984) further noted that the ME2 model produces a worse estimate as measured bigger RMSE value than zero in the extreme case where the observed matrix is given as the prior one. He argued that this

failure can only occur if there are faults in some of the assumptions about, for example, route choice and the specification of origins and destinations. As an alternative, he developed a method, called MODCOST (Modifying Origin and Destination Costs to Simulate Trips), of estimating trip matrices from traffic counts by simulating trip choices made by individuals whose different perceptions of the costs of options is represented by a random utility procedure. The method includes an iterative process between the simulation of trip choice and the correction of trips by matching them with the observed flows. This method does not use any abstract principles such as the gravity model or maximum entropy. The results from a test using data from the town of Reading showed that MODCOST achieves matrix estimates which compare favorably with other methods including the ME2 model. However, MODCOST lacks any well-established theoretical structure such as that which supports other methods.

A difficulty in the use of the ME2 model may arise when there has been a significant growth (or decline) in the total quantity of traffic between the time that the prior trip matrix was obtained and the time that the new traffic counts were made (Willumsen, 1984). Earlier, Van Zuylen (1981) recognized that ME2 gives counter-intuitive results in the case of uniform overall growth and proposed a modification to the ME2 method. In this, Van Zuylen described an alternative algorithm which allows the prior trip matrix to be uniformly scaled up or down. Bell (1983; 1984; 1985) has also developed an improved version of the ME2 model, referred to as the 'the log-linear model'. This improved version generates an estimated trip matrix which is invariant to the application of uniform scaling to the prior estimates. Bell's work is described here in more detail, because it is related to the new formulation proposed in Chapter 5.

Bell (1983) derived a model which maximizes the joint probability of observing T subject to the link flow constraints, based on the assumption that trips are multinomially distributed. The model proposed by Bell (1983) was

P4.6

$$\text{Max}_{\underline{T}} \text{So}(\underline{T}, \underline{t}) = T_{..}(\log_e T_{..} - 1) - \sum_{i,j} T_{ij}(\log_e (T_{ij}/t_{ij}) - 1 + \log_e t_{..}) \quad (4.31)$$

s.t.

$$\sum_{i,j} P_{ij}^a T_{ij} = \bar{V}_a, \quad a=1, \dots, M, \quad i=1, \dots, N, \quad j=1, \dots, N. \quad (4.32)$$

$$T_{ij} \geq 0 \quad (4.33)$$

where $T_{..} = \sum_{i,j} T_{ij}$ and $t_{..} = \sum_{i,j} t_{ij}$.

The formal solution to Problem P4.6 can be obtained by using the Lagrange method. That is

$$T_{ij} = \phi t_{ij} \prod_{a=1}^M X_a^{P_{ij}^a} \quad (4.34)$$

where $\phi = T_{..}/t_{..}$ and $X_a = \exp(-\lambda_a)$.

The formal solution (4.34) differs from that from the ME2 model (4.30) only in so far as a scale parameter ϕ is included. The final solution can be obtained by solving (4.34) together with the link flow constraints (4.32) using a Newton method. The Newton fitting procedure converges faster than the multiproportional method but requires more computer memory. Another important improvement on the ME2 model was to accommodate errors in traffic counts within the model. The model permits the calculation of asymmetric confidence intervals for the estimated trip matrix, given or assumed variances and covariances for the measurements of traffic counts.

Later, Maher (1987) also pointed out that the estimates produced by the ME2 model are biased in some cases. According to his analysis, in the case of uniform growth since the prior trip matrix, the ME2 method overestimates those O-D flows which have been counted many times relative to those which have been counted few times. Even when the growth is non-uniform, the bias is still present. These findings were confirmed algebraically by the use of small examples and numerically by the simulation of many somewhat larger examples. As another modification

to the ME2 method, Maher proposed a method of using a simple two-stage algorithm. This was made by extending Van Zuylen's approach (1981). The extended algorithm was

A4.2

(step 0) Initialize $n=0$ and set the initial prior matrix $\underline{u}^{(n)} = \underline{t}$ and $\underline{T}^{(n)} = \underline{u}^{(n)}$.

(step 1) Keep $\underline{u}^{(n)}$ fixed, perform several iterations of the multi-proportional algorithm described earlier to modify $\underline{T}^{(n)}$ and set newly modified matrix to be $\underline{T}^{(n+1)}$. If the difference between $\underline{T}^{(n)}$ and $\underline{T}^{(n+1)}$ is within the convergence limit, stop. Otherwise, move to (step 2).

(step 2) Keeping $\underline{T}^{(n+1)}$ fixed, perform several iterations of the Furness procedure (Furness, 1965) to scale up the rows and the columns of $\underline{u}^{(n)}$ so that the trip ends of $\underline{u}^{(n)}$ match those of $\underline{T}^{(n+1)}$ and set newly modified prior to be $\underline{u}^{(n+1)}$. Set $n=n+1$ and move to (step 1).

This modified approach is particularly useful because the major merit of the ME2 method, i.e. the use of the multi-proportional procedure, is still maintained. Although the convergence of the algorithm has not been proved for the general case, in a small example Maher showed that this method recovers correctly the values of \underline{T} whereas the ME2 method does not.

As an extension to the ME2 model, Willumsen (1984) suggested a model which incorporates the uncertainty in traffic counts directly into the estimation process. The model proposed by Willumsen (1984) was

P4.7

$$\text{Min } \sum_{i,j} T_{ij} (\log_e T_{ij}/t_{ij}-1) + \omega \sum_{i,j} V_a (\log_e V_a/\bar{V}_a-1) \quad (4.35)$$

s.t.

$$\sum_{i,j} T_{ij} P_{ij}^a = V_a, \quad a=1,\dots,M, \quad i=1,\dots,N, \quad j=1,\dots,N. \quad (4.36)$$

$$T_{ij} \geq 0 \quad (4.37)$$

where V_a is the flow on link a estimated by the model, and ω is the relative weight to uncertainties and errors in the traffic counts compared to uncertainties in the prior trip matrix.

Recently, following Willumsen (1984), Brenninger-Göthe, Jörnsten and Lundgren (1988) formulated the estimation problem as a multiobjective programming problem which allows the specification of different objectives depending on the beliefs in the prior trip matrix and traffic counts. According to their multiobjective programming formulation, the estimation problem can be interpreted as: one objective is to satisfy the prior trip matrix and a second objective is to satisfy the traffic counts values. However, there does not exist any clear way to specify the relative weights of these two objectives.

4.3.2.2 Information minimization approach

Van Zuylen (1978) developed a model to estimate a trip matrix from traffic counts based on the information minimization formalism. Since the information available in the traffic counts on the links is insufficient to determine a complete trip matrix, it seems reasonable to choose a trip matrix that adds as little information as possible to the knowledge contained in the link flow constraints. This approach has been followed using Brillouin's information measure.

The information contained in a set of N observations where the state k has been observed n_k times is defined by Brillouin (1956) as:

$$I = - \log_e N! \prod_k \frac{q_k^{n_k}}{n_k!} \quad (4.38)$$

where q_k is a priori probability of observing state k . If the observations are counts on a particular link, it is possible to define state ij as the state in which the vehicle observed has been travelling between origin i and destination j . So,

$$n_{ij}^a = T_{ij} P_{ij}^a \quad (4.39)$$

We can also express the a priori probability of observing state ij on link a as a function of a priori information about the trip matrix as

$$q_{ij}^a = \frac{t_{ij} P_{ij}^a}{\sum_{i,j} t_{ij} P_{ij}^a} \quad (4.40)$$

where t_{ij} is the a priori number of trips between i and j provided, for example, by an old trip matrix. The information contained in \bar{V}_a counts on link a is then

$$I_a = - \log_e \bar{V}_a! \prod_{i,j} \frac{(t_{ij} P_{ij}^a / S^a)^{T_{ij} P_{ij}^a}}{(T_{ij} P_{ij}^a)!} \quad (4.41)$$

where $S^a = \sum_{i,j} t_{ij} P_{ij}^a$.

Using Stirling's approximation, it is possible to obtain

$$I_a \cong \sum_{i,j} T_{ij} P_{ij}^a \log_e \frac{T_{ij} S^a}{\bar{V}_a t_{ij}} \quad (4.42)$$

Summing up over all the links in the network with counts, the total information contained in the observed link flows is

$$I = \sum_a \sum_{i,j} T_{ij} P_{ij}^a \log_e \frac{T_{ij} S^a}{\bar{V}_a t_{ij}} \quad (4.43)$$

The problem of finding a trip matrix consistent with the observations and adding a minimum of extra information to them is equivalent to minimizing the measure I subject to the flow constraints. That is

P4.8

$$\text{Min } I = \sum_a \sum_{i,j} T_{ij} P_{ij}^a \log_e \frac{T_{ij} S^a}{\bar{V}_a t_{ij}} \quad (4.44)$$

s.t.

$$\sum_{i,j} P_{ij}^a T_{ij} = \bar{V}_a, \quad a=1,\dots,M. \quad (4.45)$$

$$T_{ij} \geq 0 \quad (4.46)$$

The formal solution to this problem can be obtained by differentiation of the Lagrangian. Finally, we obtain

$$T_{ij} = t_{ij} \prod_{a=1}^M X_a P_{ij}^a / \sum_{k=1}^M P_{ij}^k \quad (4.47)$$

$$\text{where } X_a = \frac{V_a e^{-(1+\lambda_a)}}{\sum_{i,j} t_{ij} P_{ij}^a}$$

The optimum solution \underline{T} is obtained by solving Equations (4.47) and (4.45) simultaneously. This is again a case which can be efficiently solved by the multi-proportional method. It is interesting to see that the solution (4.47) is very similar to the solution (4.30) obtained from the ME2 model. The main difference between these two resides in the exponents, $P_{ij}^a / \sum_k P_{ij}^k$, for Van Zuylen's model and simply P_{ij}^a in Willumsen's model. This difference has the form of weights to be associated with observations on link a . The similarity is not surprising as the close relationships between entropy maximization and minimum information principles have long been recognized as, generally speaking, the state of maximum disorder is equivalent to the one containing minimum information. Van Zuylen and Willumsen (1980) has showed that Van Zuylen's minimum information model can also be derived using the entropy maximization principle.

4.3.3 Methods using statistical inference techniques

As described in Section 4.3.2, the estimation methods which use the principles of maximum entropy and minimum information assume that the prior trip matrix and traffic counts are known with certainty. In reality, prior trip matrices are subject to variations as they are obtained, for example, from old studies or sampling surveys. Also, measured traffic counts are random variables due to various errors, and neither are the modelled assignment proportions \underline{P} known with certainty. The group of estimation methods described below explicitly consider variations in the prior trip matrix and errors in measured traffic counts by using statistical inference techniques. In the end, the

estimated trip matrix is expressed as a function of the variances and covariances as well as the observed values of the prior trip matrix and traffic counts used in the estimation. However, it is still assumed that the modelled assignment proportions \underline{P} are fixed.

In this section, three different methods - Bayesian, maximum likelihood and least square methods - will be reviewed.

4.3.3.1 The Bayesian approach

As described in Section 3.3.1, the problem of estimating a trip matrix from traffic counts is underspecified: there are many trip matrices which satisfy the link flow constraints whenever these are feasible at all. In order to determine a unique solution, some other information is added in the form of the prior information (for example, a prior trip matrix). Then, the combination of prior information and traffic counts produces a posterior estimate of the trip matrix \underline{T} . In the two previous methods - maximum entropy and minimum information - a unique solution is determined by using some objective functions, generally in the form of a generalised measure of distance between the estimated trip matrix \underline{T} and the prior one \underline{t} . In these, the posterior trip matrix is an amalgam of the prior trip matrix and the observations on some links, but in this amalgam as little weight as possible is given to the prior. Furthermore, in these previous methods only point values are specified in the prior trip matrix but no measure of the degree of belief in this prior matrix is given or allowed for.

The problem of combining prior beliefs and observations to produce posterior beliefs is a standard one in Bayesian statistical inference. If one has complete confidence in one's prior beliefs, then no random observation, however remarkable, will change in one's opinions and the posterior beliefs will be identical to the prior ones. On the other hand, if one has little confidence in the prior, the observations will play the dominant role in determining the posterior beliefs. It is envisaged that the prior information may come from an old transportation study and also the observations made on the link flows will be subject to random error. The uncertainties in the prior beliefs and the

observations could be of comparable magnitude. An information minimizing approach, then, could well be throwing away useful information in the prior. Furthermore, a greater degree of confidence may be held for some parts of the prior beliefs than for others. For example, a recent transport study may have been carried out on part of the current study area and so some elements of the trip matrix may be known much more accurately than others. In this case, then, there are varying degrees of belief in different parts of the prior (Maher, 1983).

Having considered variations in the prior matrix and the observations, Maher (1983) developed a method to estimate a trip matrix from traffic counts based on Bayes' theorem which states that

$$Q(\underline{\theta}|\underline{V}) = \frac{f(\underline{V}|\underline{\theta})Q(\underline{\theta})}{\int f(\underline{V}|\underline{\theta})Q(\underline{\theta})d\underline{\theta}} \quad (4.48)$$

where $Q(\underline{\theta})$ is the prior probability density of the parameter $\underline{\theta}$, $f(\underline{V}|\underline{\theta})$ is the probability of the observations \underline{V} given the parameter values $\underline{\theta}$ and $Q(\underline{\theta}|\underline{V})$ is the posterior probability density of $\underline{\theta}$ given the observations \underline{V} .

Since the denominator in (4.48) is a constant for the prior $\underline{\theta}$, the posterior probability density of $\underline{\theta}$ is proportional to the product of the probability of the observations given the prior $\underline{\theta}$ and the prior probability density of $\underline{\theta}$:

$$Q(\underline{\theta}|\underline{V}) \propto f(\underline{V}|\underline{\theta})Q(\underline{\theta}) \quad (4.49)$$

Maher assumed the multivariate normal distributions (MVN) for the prior $\underline{\theta}$ and the random errors in the observations \underline{V} . Despite some misgivings about its use for elements with small means, the multivariate normal distribution seems to be an appropriate choice for the distribution of both the prior and the random errors in traffic counts. Following these assumptions, Maher showed that the posterior distribution $Q(\underline{\theta}|\underline{V})$ is also MVN.

Let us assume that the prior distribution of the parameter $\underline{\theta}$ is

MVN($\underline{t}, \underline{Z}_0$) where \underline{Z}_0 is dispersion matrix for $\underline{\theta}$, the distribution of the random errors $\underline{\epsilon}$ in the observations \underline{V} is MVN($0, \underline{W}$) where \underline{W} is the dispersion matrix for $\underline{\epsilon}$ and the posterior distribution for the parameter $\underline{\theta}$ is MVN($\underline{T}, \underline{Z}_1$) where \underline{Z}_1 is the dispersion matrix for the posterior estimated. Then we obtain the following updating equations for the mean and dispersion matrices of the posterior estimated (Maher, 1983):

$$\underline{T} = \underline{t} + \underline{Z}_0 \underline{P}' (\underline{W} + \underline{P} \underline{Z}_0 \underline{P}')^{-1} (\underline{V} - \underline{P} \underline{t}) \quad (4.50)$$

$$\text{and } \underline{Z}_1 = \underline{Z}_0 - \underline{Z}_0 \underline{P}' (\underline{W} + \underline{P} \underline{Z}_0 \underline{P}')^{-1} \underline{P} \underline{Z}_0 \quad (4.51)$$

The central calculation in (4.50) and (4.51) is the inversion of the symmetric matrix $(\underline{W} + \underline{P} \underline{Z}_0 \underline{P}')$. Prior beliefs are modified by observations to produce posterior beliefs: the stronger the prior beliefs the less influence the observations will have in determining the posterior beliefs. The posterior beliefs are a weighted average of the prior beliefs and the observations, and the relative weights in this average are determined by the relative magnitudes of the two dispersion matrices: \underline{W} for the observation errors and \underline{Z}_0 for the prior beliefs. The Bayesian method has the advantage of allowing more flexibility than previous methods in the degree of beliefs. Furthermore, it allows for different degrees of belief in different elements of the prior. The methods based on maximum entropy and minimum information are seen to be just extreme cases of a whole range of possibilities. Also, Maher suggested the incorporation of a sensitivity analysis in the Bayesian method so that the effects of small changes in the assumed values of the parameters on the solution can be investigated without repeating the entire estimation process.

Maher (1983) tested the Bayesian method for the problems of estimating turning flows at a junction and estimating a trip matrix from traffic counts on a network. The results from the Bayesian method were compared with those of the maximum entropy method. It was seen that there is close agreement in the solutions between two methods. The closeness of this agreement is influenced by both the prior means and the prior variances. From the computational point of view, the Bayesian method requires more storage for the inversion of the matrices but less computation time than the maximum entropy or minimum information

approaches.

4.3.3.2 The maximum likelihood method

Spiess (1987) argued against the vague use of the prior information in the matrix estimation process. According to his finding, various descriptions are used in the literature for the prior \underline{t} such as initial estimate, out-dated trip matrix, a priori guess, target matrix and reference matrix. In previous methods, the matrix \underline{t} is not essential to the formulation of the model so that it is possible to obtain an estimate of the trip matrix even in the absence of a prior matrix: that is achieved by setting $t_{ij}=1$, for all i and j . However, the results of these methods, when used without the prior matrix \underline{t} , have so far not been very convincing. It has been recognized that the accuracy of the resulting estimates is highly sensitive to the information contained in the prior trip matrix \underline{t} . For example, see Robertson (1984), Atkins (1987), and Lam and Lo (1990). Given this rather imprecise and perhaps confusing role of the prior trip matrix \underline{t} , it is not surprising that previous methods lack statistical proofs of validity for the resulting estimates.

Following this criticism, Spiess (1987) developed a method of estimating a trip matrix from a prior trip matrix when the volumes on a subset of the links of the network and/or the total productions and attractions of the zones are known. In this method, the prior trip matrix is not optional but becomes an essential part. In his model, the prior matrix \underline{t} obtained by sampling for each O-D pair is assumed to be a Poisson distributed random variable with unknown mean $\rho\underline{T}$. The coefficients ρ represent the sampling factor for O-D pair, i.e. the fraction of the population that has been observed. Thus, \underline{t} is one observation of a set of random variables that have independent Poisson distributions with means $\rho\underline{T}$. Since traffic counts are usually collected on a day-to-day basis, they are considered to be more reliable compared to the prior information and therefore it is possible to assume that their true mean values are known, or at least good approximations thereof. Later, this assumption is relaxed and the model is extended to accommodate errors in observed link flows.

Under the hypothesis that the prior matrix \underline{t} is obtained by observing an independent Poisson process with unknown mean $\rho \underline{T}$ where the sampling fractions ρ are known, the joint probability of observing the prior \underline{t} is

$$\text{Prob}(\underline{t}) = \prod_{i,j} \frac{(\rho_{ij} T_{ij})^{t_{ij}} e^{-\rho_{ij} T_{ij}}}{t_{ij}!} \quad (4.52)$$

Furthermore, we know that the population satisfies the link flow constraints. Applying the maximum likelihood estimation technique to this problem amounts to finding the estimated matrix \underline{T} which satisfies the given link flow constraints and yields the maximum probability of (4.52) for observing the prior \underline{t} . By replacing the probability (4.52) with its logarithm and discarding constant terms, the maximum likelihood model can be formulated as:

P4.9

$$\text{Max } \sum_{i,j} (t_{ij} \log_e(T_{ij}) - \rho_{ij} T_{ij}) \quad (4.53)$$

s.t.

$$\sum_{i,j} P_{ij}^a T_{ij} = \hat{V}_a, \quad a=1,\dots,M. \quad (4.54)$$

$$T_{ij} \geq 0 \quad (4.55)$$

Problem **P4.9** is convex in \underline{T} . If the link flow constraints are feasible and mutually consistent, then the existence of a global optimum solution is assured. The solution to **P4.9** may therefore be obtained by any standard solution method for convex programming problems. In practice, however, matrix estimation problems are typically of a large size. Solution methods that require access to all data simultaneously may be difficult to implement. Spiess (1987) suggested the use of the cyclic coordinate decent algorithm, which is based on the idea of successive relaxation of all but one of the link flow constraints. This solution method is similar to the multi-proportional method used to solve the maximum entropy model.

In the problem **P4.9**, it was assumed that the true values of the

observed link flows \underline{V} are known. However, in reality, this is not always possible. This leads to an extension of the model, in which we interpret the observed link flows, denoted $\{\underline{V}_a\}$, to be random samples from a Poisson distribution with unknown mean $\{\tau_a V_a\}$. The positive constants $\{\tau_a\}$ are the sampling factors used in the observation of the link flows. If the observed link flows are obtained from mutually independent surveys and independently from the sampled prior matrix \underline{t} , the maximum likelihood method yields the following model:

P4.10

$$\text{Max}_{\underline{T}} \sum_{i,j} (t_{ij} \log_e(T_{ij}) - \rho_{ij} T_{ij}) + \sum_a (\underline{V}_a \log_e V_a - \tau_a V_a) \quad (4.56)$$

s.t.

$$\sum_{i,j} P_{ij}^a T_{ij} = V_a, \quad a=1,\dots,M, \quad i=1,\dots,N, \quad j=1,\dots,N. \quad (4.57)$$

$$T_{ij} \geq 0 \quad (4.58)$$

Problem **P4.10** is also a convex programming problem which can be solved by the same solution method as used for Problem **P4.9**. As described in Section 4.3.2.1, Van Zuylen and Willumsen (1980) and later Van Zuylen and Branston (1982) also used a maximum likelihood method to estimate a set of mutually consistent link flows from inconsistent ones. They too assumed that the observed link flows are Poisson-distributed and mutually independent.

It is important to note that, in order to apply the maximum likelihood method, it is necessary to give a precise interpretation of the prior matrix \underline{t} . While other assumptions on the probability distribution on the prior \underline{t} could be possible and would lead to different maximum likelihood models, the Poisson-distributed observed prior matrix \underline{t} is perhaps most appropriate in practice.

4.3.3.3 The generalized least squares method

McNeil and Hendrickson (1985a; 1985b) and Cascetta (1984) have proposed the use of the Generalised Least Squares (GLS) as a method for estimating trip matrices from traffic counts. A major attraction of the GLS method is that it allows the combination of the prior information

with traffic counts, while taking into account the relative accuracy of these two sources of data. It can be shown that the GLS approach is formally the same as the Bayesian approach and is closely linked to the Entropy approach. For example, see Cascetta and Nguyen (1988).

Following the work by Carey et al (1981), McNeil and Hendrickson (1985a; 1985b) suggested use of the generalised least squares method to estimate trip matrices from traffic counts. The estimation problem formulated by McNeil and Hendrickson allows the uncertainty of the estimates to be forecast as well as explicitly including all available information in the form of constraints. For example, some information includes: (1) observations of some particular matrix entries, (2) observed link flows collected in a part of links in networks, (3) observations of row and column totals.

The problem formulated by McNeil and Hendrickson (1985a) was

P4.11

$$\text{Min}_{\underline{T}} (\underline{T}-\underline{t})' \underline{Z}^{-1}(\underline{T}-\underline{t}) \quad (4.59)$$

$$\text{s.t. } \underline{P}\underline{T} = \underline{V} \quad (4.60)$$

where \underline{Z} is vector used to represent weights.

Problem **P4.11** is convex and quadratic in \underline{T} and can be solved by any solution method for convex programming problems. The final solution in matrix form is given by

$$\underline{T} = \underline{t} + \underline{Z}\underline{P}'(\underline{P}\underline{Z}\underline{P}')^{-1}(\underline{V}-\underline{P}\underline{t}) \quad (4.61)$$

It can be shown that Equation (4.61) is a global optimum solution to **P4.11** if the matrix \underline{Z} is non-singular and positive definite, which is likely to be true for typical specifications of \underline{Z} .

Also, **P4.11** can be reformulated as a general constrained least squares regression problem:

P4.12

$$\text{Min } \underline{\underline{\epsilon}}' \underline{\underline{\epsilon}} \quad (4.62a)$$

s.t.

$$\underline{\underline{t}} = \underline{\underline{T}} + \underline{\underline{\epsilon}} \quad (4.62b)$$

$$\underline{\underline{PT}} = \underline{\underline{V}} \quad (4.63)$$

where $\underline{\underline{\epsilon}}$ is a vector of errors where $E(\underline{\underline{\epsilon}})=\underline{\underline{0}}$ and $\text{Var}(\underline{\underline{\epsilon}})=\sigma^2\underline{\underline{Z}}$

Solving Problem P4.12 by the least squares regression method leads to the same solution as Equation (4.61). In this case, the solution (4.61) can be interpreted as the best linear unbiased estimate of $\underline{\underline{T}}$ for the problem P4.12. Furthermore, the variance-covariance matrix of the estimated matrix $\underline{\underline{T}}$ is given by

$$\text{Var}(\underline{\underline{T}}) = \sigma^2[\underline{\underline{Z}} - \underline{\underline{Z}}\underline{\underline{P}}'(\underline{\underline{P}}\underline{\underline{Z}}\underline{\underline{P}}')^{-1}\underline{\underline{P}}\underline{\underline{Z}}] \quad (4.64)$$

By making some distributional assumptions about the error terms, confidence intervals for the estimate $\underline{\underline{T}}$ can be calculated. Using a simple example, McNeil and Hendrickson (1985a) compared the results from the quadratic programming method with the results from other methods including the maximum entropy and information minimization methods. It was found that in these examples these methods yield similar numerical results.

Apart from McNeil and Hendrickson (1984a; 1985b), Cascetta (1984) also developed a generalized least squares model to estimate a trip matrix from traffic counts. The estimator resulting from the model is to combine the prior estimate, possibly obtained by a direct sample survey or demand model, with traffic counts. The model explicitly considers measurement errors in link flows, misspecification errors in the traffic assignment, and variations in the prior matrix. Two cases were considered: a more general one, in which the estimator was stochastically constrained to the observed flows considered to be random variables, and another in which the estimator is deterministically constrained to the observed flows. The latter one with the deterministic observed flows results in the same quadratic programming formulation as

the one proposed by McNeil and Hendrickson (1985a). We here review only the general case.

If \underline{t} is the prior estimate obtained by either direct survey or demand models, it can be expressed as

$$\underline{t} = \underline{T} + \underline{\varepsilon} \quad (4.65)$$

where $\underline{\varepsilon}$ is a vector of random errors with mean $\underline{\mu}$ and dispersion matrix \underline{Z} .

Also, if \underline{V} is the vector of observed link flows, it can be posed

$$\underline{V} = \underline{P}\underline{T} + \underline{\eta} \quad (4.66)$$

where $\underline{\eta}$ is a vector of random errors with mean $\underline{\delta}$ and dispersion matrix \underline{W} .

Combining two linear equations (4.65) and (4.66) leads to a linear equation:

$$\begin{bmatrix} \underline{t} \\ \underline{V} \end{bmatrix} = \begin{bmatrix} \underline{I} \\ \underline{P} \end{bmatrix} \underline{T} + \begin{bmatrix} \underline{\varepsilon} \\ \underline{\eta} \end{bmatrix} \quad (4.67)$$

where \underline{I} is an identity matrix.

The minimum variance estimator or the generalized least square estimator \underline{T}^* from Equation (4.67) can be found by solving the following quadratic programming problem:

P4.13

$$\text{Min}_{\underline{T}} \begin{bmatrix} \underline{t} - \underline{T} \\ \underline{V} - \underline{P}\underline{T} \end{bmatrix}' \begin{bmatrix} \underline{Z}^{-1} \underline{0} \\ \underline{0} \quad \underline{W}^{-1} \end{bmatrix} \begin{bmatrix} \underline{t} - \underline{T} \\ \underline{V} - \underline{P}\underline{T} \end{bmatrix} \quad (4.68)$$

$$\text{s . t. } \underline{T} \geq \underline{0} \quad (4.69)$$

Under the hypothesis of inactive inequality constraints, the generalized least square estimator \underline{T}^* is obtained by equating to zero the first partial derivatives of (4.68) with respect \underline{T} . The result is:

$$\underline{T}^* = (\underline{Z}^{-1} + \underline{P}'\underline{W}^{-1}\underline{P})^{-1}(\underline{Z}^{-1}\underline{t} + \underline{P}'\underline{W}^{-1}\underline{V}) \quad (4.70)$$

It can be shown that the mean of \underline{T}^* is:

$$\underline{E}(\underline{T}^*) = \underline{T} + (\underline{Z}^{-1} + \underline{P}'\underline{W}^{-1}\underline{P})^{-1}(\underline{Z}^{-1}\underline{\mu} + \underline{P}'\underline{W}^{-1}\underline{\delta}) \quad (4.71)$$

and its dispersion matrix is:

$$\underline{D}(\underline{T}^*) = (\underline{Z}^{-1} + \underline{P}'\underline{W}^{-1}\underline{P})^{-1} \quad (4.72)$$

It can be seen that there are strong formal similarities between the Bayesian estimators and the generalized least squares estimators (Cascetta, 1984; Cascetta and Nguyen, 1988). The main difference is that the Bayesian approach is based on the probability distribution - the multivariate normal distribution - on the prior matrix and traffic counts whereas the generalised least squares approach is based on the starting dispersion matrices without any distributional assumptions on the observations.

Cascetta (1984) carried out a small simulation exercise to get a rough idea about the effects of substituting different approximate dispersion matrices to the true ones and to compare the performances of the proposed estimator with those of the maximum entropy one. It resulted that the substitution of an estimated dispersion matrix in place of the true one produced only a slight worsening of the estimator's characteristics and all the generalized least squares estimators considered had a mean square error lower than the maximum entropy estimator.

Recently, Bell (1991) proposed a method for solving a generalised least squares problem with inequality constraints, $\underline{T} \geq \underline{c}$ where \underline{c} is a vector of non-negative constants. The proposed method replaces earlier ad hoc approaches used by Cascetta (1984). According to Bell's analysis,

unless inequality constraints are introduced into the GLS problem, the GLS estimates of some smaller O-D movements can be negative, violating the constraints. Using small examples, it was shown that the proposed solution method solves the problem more satisfactorily than ad hoc approaches.

4.4. Methods using equilibrium traffic assignment

The estimation methods reviewed in the preceding sections are based on the assumption that it is possible to obtain the route choice proportions \underline{P} independently from the trip matrix estimation process. This is only possible for situations in which proportional assignment methods are considered to be sufficiently realistic. Whenever congestion in networks plays an important role in route choice, this assumption becomes unrealistic: the calculation of the assignment proportions \underline{P} and the estimation of the trip matrix \underline{T} become interdependent.

A number of methods have been proposed for the problem of estimating trip matrices from observed link flows when congestion effects are taken into account. They can be classified into three methods: Willumsen's method, Nguyen's method, and Fisk and Boyce's method. In these methods, the equilibrium traffic assignment satisfying Wardrop's first principle is used, since it is the most preferred choice for the traffic assignment in congested networks because of its practical and theoretical advantages. This section reviews each of these three estimation methods in detail.

4.4.1 Willumsen's method

As an extension of the ME2 model into congested conditions, Willumsen proposed a use of a sequential solution method of alternately performing trip matrix estimation and equilibrium assignment (Hall, Van Vliet and Willumsen, 1980; Willumsen, 1981a; 1982). This heuristic method was originally used by Holm et al (1976) for calibrating traffic demand models using traffic counts. It is intended ultimately to find mutually consistent equilibrium assignment proportions which are in turn used to achieve a trip matrix consistent with observed traffic counts. The proposed solution method (Willumsen, 1981a) includes the following sequential steps:

A4.3

- (step 1) Assign the prior trip matrix \underline{t} using equilibrium assignment to obtain the route choice proportions $\underline{P}^{(0)}$ and set $n=1$.
- (step 2) Estimate $\underline{T}^{(n)}$ using $\underline{P}^{(n-1)}$ and the observed link flows \underline{V} by the ME2 model.
- (step 3) Assign $\underline{T}^{(n)}$ using the equilibrium assignment and obtain the new route choice proportions $\underline{P}^{(n)}$.
- (step 4) Set $n=n+1$. Return to (step 2) unless the changes in $\underline{P}^{(n)}$ or $\underline{T}^{(n)}$ have been sufficiently small.

There are two main difficulties with this sequential solution method (Willumsen, 1981a; 1982). The first is that convergence is not guaranteed. The sequential method assumes fixed demand during the equilibrium assignment in (step 1) and (step 3), and fixed route choice proportions during the trip matrix estimation in (step 2). In reality, route choice proportions will vary with the demand due to congestion. Because of this, the method does not converge in all cases.

The first difficulty may be illustrated by using a simple example.

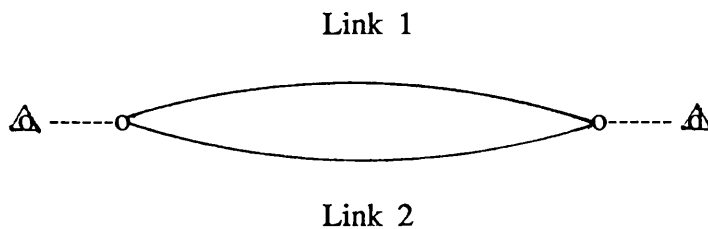


Figure 4.1 A simple example network used for testing of the convergence of the sequential solution method

As shown in Figure 4.1, the example network used has one origin o , one destination d and two links 1 and 2. We shall assume that the equilibrium flows on links 1 and 2 can be calculated directly from the

following explicit relationships without performing the equilibrium assignment.

$$V_1^* = T_{od}, \quad \text{if } T_{od} \leq 1200 \quad (4.73)$$

$$= 0.8T_{od} + 240, \quad \text{if } T_{od} > 1200$$

$$V_2^* = 0, \quad \text{if } T_{od} \leq 1200 \quad (4.74)$$

$$= 0.2T_{od} - 240, \quad \text{if } T_{od} > 1200$$

Suppose that the counts value of the prior trip matrix t_{od} is 1600 vehicles and that a single traffic count $\bar{V}_2=100$ vehicles is available on link 2. In this case, we can calculate the solution $T_{od}^*=1700$ analytically, which gives assigned flows V_2^* on link 2 which are equal to 100. However, here we shall consider the estimation of a trip matrix by the sequential method. The algorithm A4.3 of the sequential method described above estimates a trip matrix in the following way. If the prior trip matrix $T_{od}^{(0)}=t_{od}=1600$ is assigned to the network, the equilibrium link flows $V_2^{*(0)}=80$ and the route choice proportion $P_{od}^1=0.05$ are calculated using the relationships (4.73) and (4.74). From the link flow constraint $P_{od}^2 T_{od} = \bar{V}_2$, we can estimate the trip matrix $T_{od}^{(1)}=2000$. If $T_{od}^{(1)}$ is now assigned, $V_2^{*(1)}=160$ and $P_{od}^2=0.08$ are calculated using the relationships (4.73) and (4.74). This gives the estimated trip matrix $T_{od}^{(2)}=1250$ from the link flow constraint $P_{od}^2 T_{od} = \bar{V}_2$. By repeating the same process, the results shown in Table 4.4 are obtained.

Table 4.4 Estimating a trip matrix using the simple example by the sequential solution method

n	$P_{od}^2(T_{od}^{(n-1)})$	$T_{od}^{(n)}$	$V_2^{*(n)}$	$ V_2^{*(n)} - \bar{V}_2 $	$ T_{od}^* - T_{od}^{(n)} $
0	-	1600	80	20	100
1	0.05	2000	160	60	300
2	0.08	1250	10	90	450
3	0.01	12500	2260	2160	10800
⋮	⋮	⋮	⋮	⋮	⋮

As shown in Table 4.4, the sequential method fails to converge to a mutually-consistent solution. Rather, it appears to oscillate divergently.

We consider a sequential scheme for determining the self-consistent solutions of two mutually-dependent sub-problems

$$y = f(x) \text{ and } x = h(y) \quad (4.75a)$$

where f and h express each of the two mutually-dependent sub-problems.

The sequential method of solving the problems (4.75a) can be written in the equivalent form

$$\begin{aligned} x_{n+1} &= h[f(x_n)], \quad n=0,1,\dots \\ &= g(x_n) \end{aligned} \quad (4.75b)$$

where $g(x_n)$ is the composite function $h[f(x_n)]$.

The convergence of the sequential scheme (4.75b) can be guaranteed by the Lipschitz condition (Isaacson and Keller, 1966, pp 85-91). The Lipschitz condition states that the sequential scheme (4.75b) converges provided that for all x in a neighborhood for the solution,

$$\left| \frac{\partial g}{\partial x} \right| \leq L \quad (4.76)$$

for some constant L in the interval $[0,1)$.

Applying the Lipschitz condition to the simple example shown in Figure 4.1, we shall derive a convergence condition for the sequential method which applied to this simple network. Let $T_{od}^{(\eta)}$ be the trip matrix estimated at the iteration n . Let $V_2^{*(n)}$ be the equilibrium link flows at the iteration n . Let the equilibrium link flows $V_2^{*(n)}$ of $T_{od}^{(\eta)}$ be calculated from the specific explicit assignment relationship

$$V_2^*(T_{od}) = aT_{od} + b \quad (4.77)$$

Also, let $P_{od}^2(T_{od}^{(\eta)})$ be the route choice proportion of $T_{od}^{(\eta)}$ using the link 2. Then, we can obtain the following recursive relationship from the sequential matrix estimation procedure.

$$T_{od}^{(\eta+1)} = \frac{\hat{V}_2}{P_{od}^2(T_{od}^{(\eta)})} \quad (4.78)$$

$$= \frac{\hat{V}_2}{V_2^{*(n)}} T_{od}^{(\eta)} \quad (4.79)$$

Then, the Lipschitz condition (4.76) indicates that the sequential algorithm will converge provided that

$$\left| \frac{\partial}{\partial T_{od}} \left(\frac{\bar{V}_2 T_{od}}{V_2^*(T_{od})} \right) \right| < 1. \quad (4.80)$$

Using Equation (4.77) and taking the derivative, we obtain

$$\frac{\partial}{\partial T_{od}} \left(\frac{\bar{V}_2 T_{od}}{V_2^*(T_{od})} \right) = \frac{\bar{V}_2(aT_{od}+b-aT_{od})}{(aT_{od}+b)^2} \quad (4.81)$$

Using the approximation $aT_{od}+b \cong \bar{V}_2$, we have

$$\frac{\partial}{\partial T_{od}} \left(\frac{\bar{V}_2 T_{od}}{V_2^*(T_{od})} \right) \cong \frac{b}{aT_{od}+b} \quad (4.82)$$

Finally, substituting (4.82) into (4.80), we obtain the following result:

'The sequential method converges if $|b| < aT_{od}+b$. More specifically, if $b \geq -\bar{V}_2$, the sequential method will converge. Otherwise, if $b < -\bar{V}_2$, it can diverge.'

If this result is applied to the simple example, because $b = -240 < -100 = -\bar{V}_2$, the sequential method need not converge starting with the initial solution $T_{od}^{(0)} = 1600$. Thus, the use of the Lipschitz condition provides a theoretical explanation for the observed divergence of the sequential method in some cases.

From the results obtained above, a practical suggestion can be made for link observations when estimating trip matrices from observed link flows. The links on a network can be classified into two groups. The links which are used at the initial route choice are of the first group. The links of the first group will have positive link flows in the initial all or nothing assignment. The links which are not classified as the first group are of the second group. Thus, the links of the second group do not carry any flow in the initial all or nothing assignment. For example, in the case of the simple example, if the flows in link 1,

which is of the first group, are observed, the sequential method will converge, provided that an initial solution close to the final solution is given. In general, this result can be applied for making observations of the links when estimating trip matrices and so it will help the sequential method to converge.

The second difficulty is that, as described in Section 3.3.4, the route choice proportions \underline{P} are in general not uniquely determined by the equilibrium assignment process. The sequential method extracts route choice proportions from the trees and optimum flow combination parameters λ in the Frank-Wolfe algorithm described in Section 3.2.2. This is only an ad hoc device chosen to explore ways of extending the ME2 model to equilibrium assignment conditions. The method used is to set:

$$\underline{P}^{(n+1)} = (1-\lambda^{(n)})\underline{P}^{(n)} + \lambda^{(n)}\underline{P}^*(\underline{c}^{(n)}) \quad (4.83)$$

where $\underline{P}^*(\underline{c}^{(n)})$ is the route choice proportions by all or nothing assignment using link costs $\underline{c}^{(n)}$, $\underline{P}^{(n)}$ is the route choice proportions up to iteration n , and $\lambda^{(n)}$ is the optimum linear combination parameter from the Frank-Wolfe equilibrium assignment at iteration n .

Willumsen (1981a; 1982) tested this sequential method using the Reading data which had already been used for the validation of the ME2 model. The test was carried out using Tuesday 19 October data base with the extended set of counts (trip end counts included). The two hour traffic counts (16:10 to 18:10) were appropriately scaled when used to update costs with the cost-flow relationships. First of all, Willumsen tested how well the equilibrium assignment programs used reproduce the observed flows in the Reading area. In order to do this, the observed trip matrix was loaded by each of all or nothing and equilibrium and the resulting flows were compared with the observed ones. As shown in Table 4.5, equilibrium assignment produces link flows which are closer to the observed ones than all or nothing assignment. It was also observed that all or nothing assignment produces reasonable flow levels suggesting that in this case it is not a bad approximation.

Table 4.5 Observed vs loaded flows from 19 October trip matrices
(extracted from Willumsen (1981a))

I n d i c a t o r	Loading Technique	
	All or nothing	Equilibrium assignment
RMSE	18.8	16.0
% RMSE	20	16

Next, the new trip matrix was estimated by the sequential solution method of the ME2 model and this estimate was compared with the observed trip matrix. As goodness-of-fit statistics, the root mean square error (RMSE) and the percent root mean square error (%RMSE) between the estimated and the observed were used for comparing two different trip matrices. From the final results, as shown in Table 4.6, the following comments on the sequential method were made (Willumsen, 1981a):

- (1) The sequential method seems to produce an improved estimate of the sampled trip matrix albeit at a high cost in computing time. This improvement is, however, not very large.
- (2) The greatest improvement seems to be produced during the first cycle. Additional iterations did not improve the estimated matrix much.
- (3) The method is not completely 'well behaved', in the sense that after certain iterations the goodness of fit worsens, albeit only marginally.
- (4) Although not apparent in Table 4.6, it was observed that full convergence of the ME2 model was not achieved because the flow constraints were not mutually consistent. The sequential method tends to generate certain route choice proportions which are not fully consistent with the observed link flows.
- (5) On the whole it cannot be said that the sequential method is entirely satisfactorily.

Table 4.6 Tests with the sequential method for estimating trip matrices
(extracted from Willumsen (1981a))

Indicator	1st cycle					2nd cycle			3rd cycle		
	iteration					iteration			iteration		
	1	2	3	4	5	2	3	4	2	3	4
RMSE	2.1	2.1	2.0	2.0	2.0	2.0	2.0	2.1	2.0	2.0	2.0
%RMSE	169	167	163	164	162	162	160	179	160	161	160

Note (1) Each cycle is an equilibrium assignment.

(2) Each iteration is a Frank-Wolfe iteration within the equilibrium assignment.

(3) Trip matrices are updated by the ME2 process after each iteration of the assignment.

This sequential method of the ME2 model to estimate a trip matrix from traffic counts under congested conditions has been adopted for use with the SATURN suite (Hall, Van Vliet and Willumsen, 1980). SATURN - Simulation and Assignment of Traffic in Urban Road Networks - is a simulation-assignment model for the evaluation of traffic management schemes developed at the Institute for Transport Studies, University of Leeds. The simulation part of the SATURN model treats junctions in great detail, thus providing a good representation of the way in which equilibrium might be achieved in urban areas.

The procedure adopted in SATURN is shown in Figure 4.2 and it follows the sequential method which performs equilibrium assignment and trip matrix estimation alternately. Moreover, the assignment can be either a single equilibrium assignment with fixed flow-delay curves as represented by 'Inner Loop' or a full run of simulation and equilibrium assignment as represented by 'Outer Loop'. The paths represented by both 'Inner' and 'Outer' loops can be viewed as almost interchangeable, and a complete run of the model could pass through each several times.

This heuristic procedure adopted in SATURN was applied to a project of evaluating a number of alternative traffic management schemes for the town of Harrogate in North Yorkshire (Hall, Van Vliet and Willumsen, 1980). The procedure was used to update an old trip matrix from traffic counts. It was shown that the trip matrix updated through the SATURN

procedure halves the mean absolute difference between the assigned and observed flows by comparison with those of the old trip matrix. It was suggested that the matrix updating method implemented in SATURN has led to a substantial improvement in the goodness-of-fit between observed and predicted flows and are a valuable addition to the modelling of limited area traffic management schemes.

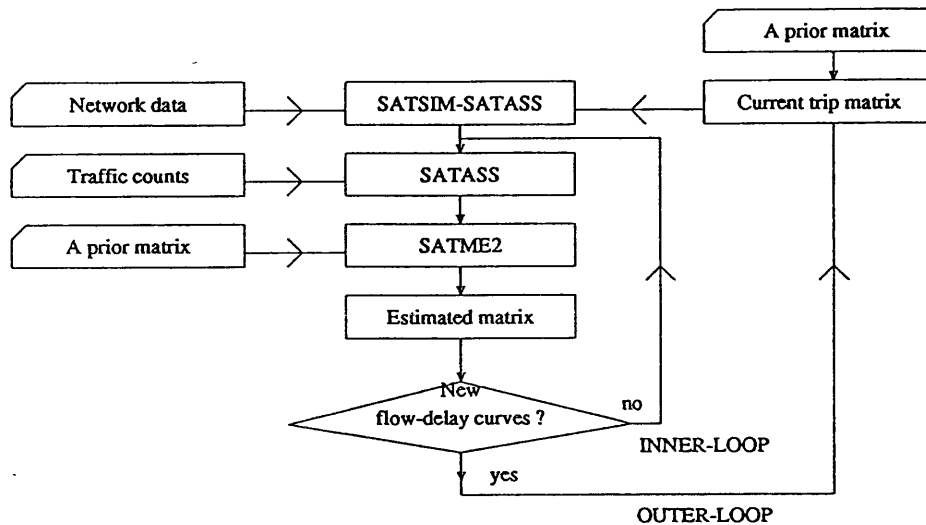


Figure 4.2 Estimating a trip matrix by the sequential ME2 method in the SATURN (taken from Van Vliet (1987))

Recently, Fisk (1988) proposed a formulation which solves the two sub-problems of equilibrium assignment and maximum entropy matrix estimation simultaneously. Fisk incorporated the equilibrium conditions as a constraint in the entropy maximizing problem by adopting the form of the variational inequality used originally by Smith (1979). The proposed formulation was:

P4.14

$$\text{Max}_{\underline{T}} S_1(\underline{T}, \underline{t}) = -\sum_{i,j} T_{ij} (\log_e T_{ij}/t_{ij} - 1) \quad (4.84)$$

$$\text{s.t.} \quad \sum_{r \in P_{ij}} \delta_{ar} h_r = \bar{V}_a, \quad a=1, \dots, M, \quad i=1, \dots, N, \quad j=1, \dots, N. \quad (4.85)$$

$$\sum_{r \in P_{ij}} h_r = T_{ij}, \quad (4.86)$$

$$\underline{c}(\underline{h}) \cdot (\underline{g} - \underline{h}) \geq 0, \quad \text{for all } \underline{g} \quad (4.87)$$

$$h_r, T_{ij} \geq 0, \quad (4.88)$$

where \underline{h} is the vector of path flows

$\underline{c}(\underline{h})$ is the vector of costs associated with path flow \underline{h}

\underline{g} is a vector of feasible path flows

P_{ij} is a set of paths through the network between i and j

This is an infinitely constrained problem generally known as the Fritz-John problem (Fisk, 1988). Solution procedures for problems with a similar structure have been investigated by Marcotte (1983) in connection with equilibrium network design problems. Following the approach of Fisk (1984b), other formulations of **P4.14** can be obtained by expressing the constraint (4.87) in other equivalent ways. This is important for solution purposes because it broadens the range of solution procedures. If $\underline{c}(\underline{h})$ is a monotone function, i.e.

$$[\underline{c}(\underline{h}) - \underline{c}(\underline{g})] \cdot (\underline{h} - \underline{g}) \geq 0 \quad (4.89)$$

$$\text{and } W(\underline{h}) = \text{Min}_{\underline{g}} \underline{c}(\underline{h}) \cdot (\underline{g} - \underline{h}) \quad (4.90)$$

then the constraint (4.87) is equivalent (Smith, 1979) to

$$W(\underline{h}) = 0 \quad (4.91)$$

Thus, for monotone path cost functions, **P4.14** is equivalent to

P4.15

$$\text{Max}_{\underline{T}} S_1(\underline{T}, \underline{t}) \quad (4.92)$$

$$\text{s.t. } \sum_{r \in P_{ij}} \delta_{ra} h_r = \bar{V}_a, \quad a=1, \dots, M, \quad i=1, \dots, N, \quad j=1, \dots, N. \quad (4.93)$$

$$W(\underline{h}) = 0 \quad (4.94)$$

$$\sum_{r \in P_{ij}} h_r = T_{ij} \quad (4.95)$$

As discussed in Fisk (1984a), $W(\underline{h})$ has the properties of a penalty function; in particular for any \underline{h} , $W(\underline{h}) \leq 0$ and $W(\underline{h}) = 0$ iff \underline{h} solves the constraint (4.87). Then, **P4.15** has the approximate penalty formulation:

P4.16

$$\text{Max}_{\underline{T}} S_1(\underline{T}, \underline{t}) + \mu W(\underline{h}) \quad (4.96)$$

$$\text{s.t.} \quad \sum_{r \in P_{ij}} \delta_{rah} h_r = \bar{V}_a, \quad a=1, \dots, M, \quad i=1, \dots, N, \quad j=1, \dots, N. \quad (4.97)$$

$$\sum_{r \in P_{ij}} h_r = T_{ij} \quad (4.98)$$

where μ is a positive constant.

Since $S_1(\underline{T}, \underline{t})$ and $W(\underline{h})$ are separable functions of \underline{T} and \underline{h} respectively and for any \underline{T} it is possible to find \underline{h} for which $W(\underline{h})=0$, this latter condition will hold at the solution for any value of μ . In other words, the sequential approach of **P4.16** approximates a solution to **P4.15**. Replacing $W(\underline{h})$ by Equation (4.90), **P4.16** takes the form of a max-min problem: possible solution algorithms for problems of this kind are given in Fisk (1984b).

However, Fisk's formulation retains the flow constraints (4.97) which associate the observed link flows with estimated path flows. While these constraints remain, there might be no feasible solution due to inconsistencies in the observed flows. Fisk did not provide any explanation of how to deal with these constraints.

4.4.2 Nguyen's method

In an equilibrium based approach to the problem of estimating a trip matrix from traffic counts, Nguyen (1977) proposed two mathematical models whose solutions are trip matrices which satisfy equilibrium assignment conditions and are consistent with observed link flows. The main advantage of these models is that they estimate trip matrices from traffic counts without using the route choice proportions \underline{P} which are unknown and not well defined in equilibrium assignment.

The first method proposed by Nguyen uses traffic counts for all links of the network. The proposed method was as follows. Let $\{\bar{V}_a\}$ denote the observed flows on link a and $\{\bar{C}_{ij}\}$ the travel costs on all used routes between origin i and destination j . For each O-D pair there

is one value of \bar{C}_{ij} since $\{\bar{V}_a\}$ is assumed to be in the equilibrium state. $\{\bar{C}_{ij}\}$ may be computed on the network by determining a shortest route between i and j : the link costs being $\{c_a(\bar{V}_a)\}$. If $\{T_{ij}\}$ is the trip matrix corresponding to the observed flows $\{\bar{V}_a\}$, then the equilibrium state is expressed by the following equality:

$$\sum_{i,j} \bar{C}_{ij} T_{ij} = \sum_a c_a(\bar{V}_a) \bar{V}_a \quad (4.99)$$

Hence necessary conditions for an estimated trip matrix $\{T_{ij}\}$ to be equal to $\{\bar{T}_{ij}\}$ are:

$$\sum_{i,j} \bar{C}_{ij} T_{ij} = \sum_a c_a(\bar{V}_a) \bar{V}_a \quad (4.100)$$

and

$$C_{ij} = \bar{C}_{ij}, \text{ for all } i\text{-}j \text{ pairs} \quad (4.101)$$

where $\{C_{ij}\}$ is the modelled cost on all used routes between i and j when $\{T_{ij}\}$ is assigned onto the network by the equilibrium assignment.

A trip matrix T satisfying (4.100) and (4.101) reproduces the observed interzonal travel cost $\{\bar{C}_{ij}\}$, the total network costs $\sum_a c_a(\bar{V}_a) \bar{V}_a$ and the observed link flows $\{\bar{V}_a\}$. Nguyen showed that a trip matrix satisfying conditions (4.100) and (4.101) can be obtained by solving the following problem:

P4.17

$$\text{Min}_{\underline{T}} Z(\underline{V}) = \sum_a \int_0^{V_a} c_a(x) dx \quad (4.102)$$

s.t.

$$\sum_{r \in P_{ij}} \delta_{ar} h_r = V_a, \quad i=1, \dots, N, j=1, \dots, N. \quad (4.103)$$

$$\sum_{r \in P_{ij}} h_r = T_{ij}, \quad (4.104)$$

$$h_r, T_{ij} \geq 0, \quad (4.105)$$

$$\sum_{i,j} \bar{C}_{ij} T_{ij} = \sum_a c_a(\bar{V}_a) \bar{V}_a \quad (4.106)$$

where h_r is flow on route r between i and j and $\delta_{ar}=1$ if h_r uses link a , $\delta_{ar}=0$ otherwise.

Using Kuhn-Tucker conditions, Nguyen proved the equivalence between P4.17 and the equilibrium conditions (4.100) and (4.101). Thus, the trip matrix obtained by solving the problem P4.17 reproduces the observed link flows $\{\hat{V}_a\}$.

As a solution procedure to solve the problem P4.17, Nguyen suggested an application of the Frank-Wolfe algorithm. The solution procedure proposed was:

A4.4

(Step 1) Select an initial feasible $\{T_{ij}\}$, for example $T_{ij} = K / \sum_o \sum_d \hat{C}_{od}$ where $K = \sum_a c_a(\hat{V}_a) \hat{V}_a$. Determine an initial flow pattern $\{V_a\}$ using $\{T_{ij}\}$.

(Step 2) Determine a shortest route between each O-D pair and let C_{ij} be the travel cost on this route.

(Step 3) Find the O-D pair 'rs' for which $C_{rs} / \hat{C}_{rs} = \text{Min}_{i,j} C_{ij} / \hat{C}_{ij}$ and load 'K/ \hat{C}_{rs} ' trips onto the shortest route from r to s. Let $\{\hat{V}_a\}$ be the resulting link flows.

(Step 4) If $|\sum_a c_a(V_a)(\hat{V}_a - V_a) / Z(V_a)| \leq \epsilon$ for a suitably chosen convergence parameter ϵ , terminate. Otherwise, continue to (Step 5).

(Step 5) Find an optimum combination λ minimizing $Z((1-\lambda)V_a + \lambda\hat{V}_a)$ subject to $0 \leq \lambda \leq 1$. (4.107)

(Step 6) Revise the trip matrix and link flows as follows:

$$T_{ij} = (1-\lambda)T_{ij} \quad \text{for all } ij \neq rs \quad (4.108)$$

$$T_{rs} = (1-\lambda)T_{rs} + \lambda \left(\frac{K}{\hat{C}_{rs}} \right) \quad (4.109)$$

$$V_a = (1-\lambda)V_a + \lambda\hat{V}_a \quad (4.110)$$

Return to (Step 2).

The first model proposed by Nguyen is only suitable for small networks where observed link flows for all links are easily obtained. Furthermore, the solution procedure may not be efficient for a large

number of O-D pairs. These considerations led to the development of the second model which has reduced the input requirements. The input data required for the second model is only a set of the interzonal travel costs $\{C_{ij}\}$. The proposed formulation was:

P4.18

$$\text{Min}_{\underline{T}} Z_1(\underline{V}, \underline{T}) = \sum_a \int_0^{V_a} c_a(x) dx - \sum_{i,j} C_{ij} T_{ij} \quad (4.111)$$

s.t.

$$\sum_{r \in P_{ij}} \delta_{ra} h_r = V_a, \quad \text{for all } i\text{-}j \text{ pairs} \quad (4.112)$$

$$\sum_{r \in P_{ij}} h_r = T_{ij} \quad (4.113)$$

$$h_r, T_{ij} \geq 0 \quad (4.114)$$

Problem **P4.18** reflects the situation in which the observed link flows constitute an equilibrium. For the special case in which no congestion effects are present, $c_a(x)$ is constant and the first term of the objective function becomes simply $\sum_a c_a V_a$. This results in a linear programming problem. Using the associated Kuhn-Tucker conditions, Nguyen proved that the optimal solution of problem **P4.18** satisfies the equilibrium conditions (4.100) and (4.101). Furthermore, it was shown that this solution has a unique set of link flows. However, Nguyen pointed out that the solution to Problem **P4.18** is in general not unique and there may be more than one trip matrix which could produce the same equilibrium link flows.

The problem **P4.18** has the same form as the formulation of the equilibrium traffic assignment problem with elastic demand. This structural similarity allows application of any equilibrium procedure for traffic assignment with elastic demand to solve the problem **P4.18**, such as the algorithm stated by Nguyen (1976).

Nguyen tested both formulations **P4.17** and **P4.18** on a small synthetic network with 4 zones and 19 one-way links. A known observed trip matrix was used to generate a set of the observed link flows. It was shown that trip matrices estimated from both methods closely reproduce the observed trip matrix and the observed link flows.

As a way of choosing a single trip matrix amongst the multiple optima for Problem P4.18, Jörnsten and Nguyen (1979) proposed an approach which combines the maximum entropy objective function with Problem P4.18. Let Ω be the set of all optimal trip matrices of Problem P4.18. Then, the most likely trip matrix consistent with the observed link flows \bar{V} is that trip matrix which solves

P4.19

$$\text{Max } -\sum_{i,j} T_{ij} \log_e T_{ij} \quad (4.115)$$

s.t.

$$\sum_{i,j} T_{ij} = T_{..} \quad (4.116)$$

$$\underline{T} \in \Omega \quad (4.117)$$

Since the set of candidates Ω is not known explicitly, an immediate approach for solving Problem P4.19 would be to approximate Ω by a polyhedron defined by a finite number of linear constraints. Accordingly, as a solution procedure for Problem P4.19, Jörnsten and Nguyen suggested a decomposition-relaxation approach. The resulting computational scheme consists essentially of postoptimising at each iteration a multiproportional problem and an equilibrium assignment problem. More precisely at a general iteration 'n', the steps are:

A4.5

(step 1) Solve the following multiproportional problem for $\underline{T}^{(n)}$,

P4.20

$$\text{Max } -\sum_{i,j} T_{ij} \log_e T_{ij} \quad (4.118)$$

s.t.

$$\sum_{i,j} T_{ij} = T_{..} \quad (4.119)$$

$$\sum_{i,j} \bar{C}_{ij} T_{ij} = C_0 \quad (4.120)$$

$$\sum_{i,j} C_{ij}^{(n)} T_{ij} \leq C_n \quad (4.121)$$

$$T_{ij} \geq 0 \quad (4.122)$$

(step 2) Determine the equilibrium link flows $\underline{V}^{(n)}$ and the minimum route travel costs $C_{ij}^{(n)}$ for the current trip matrix $\underline{T}^{(n)}$ and calculate the new constraint

$$\sum_{i,j} C_{ij}^{(n+1)} T_{ij} \leq C_{n+1} \quad (4.123)$$

where $C_0 = \sum_a C_a(\underline{V}_a) \underline{V}_a$ and $C_n = \sum_a C_a(\underline{V}_a^{(n)}) \underline{V}_a$

If the resulting total network travel cost C_n in (step 2) is equal to the observed total cost C_0 , terminate.

Another way of choosing a trip matrix of multiple optimum trip matrices produced from Nguyen's second approach was suggested by Gur, Turnquist, Schneider, LeBlanc and Kurth (1979), and Turnquist and Gur (1979). The method uses an externally generated trip matrix - the target trip matrix - to provide information on the structure of the unknown trip matrix. Gur et al (1979) and Turnquist and Gur (1979) proposed the problem of using a least squares objective function for finding the trip matrix reproducing observed link flows as closely as possible and lying closest to the target trip matrix. That approach involves solving Nguyen's second problem with a heuristic variation of the Frank-Wolfe technique. The variation is in the construction of search directions, which are changed so that the estimated trip matrix at each iteration lies close to the target trip matrix. The solution algorithm (Turnquist and Gur, 1979) includes the following steps:

A4.6

(step 1) Specify an initial trip matrix $\underline{T}^{(1)}$ and a flow-delay function for each link.

(step 2) Assign $\underline{T}^{(1)}$ to the unloaded network by using free flow conditions to obtain a set of modelled link flows $\underline{V}^{(1)}$. Denote this current solution $(\underline{V}^{(1)}, \underline{T}^{(1)})$.

(step 3) Set $n=1$.

- (step 4) Determine link costs at the current flow $\underline{V}^{(n)}$ and again build the shortest routes. Denote the resulting interzonal travel costs $\underline{C}^{(n)}$.
- (step 5) Given $\underline{T}^{(n)}$, \underline{C} and $\underline{C}^{(n)}$, find a correction trip matrix \underline{T}^* that is closer to a solution.
- (step 6) Assign \underline{T}^* to the routes built in (step 4) to obtain the correction link flows \underline{V}^* .
- (step 7) Find a optimum combination λ such that $0 \leq \lambda \leq 1$ and the new solutions $(\underline{V}^{(n+1)}, \underline{T}^{(n+1)}) = \lambda(\underline{V}^*, \underline{T}^*) + (1-\lambda)(\underline{V}^{(n)}, \underline{T}^{(n)})$ minimizing the objective function $Z_1(\underline{V}, \underline{T})$ of Problem P4.18.
- (step 8) Check the convergence criterion. If it is met, stop; otherwise, set $n=n+1$ and return to (step 4).

Within the basic framework of this algorithm, there are a number of opportunities for variation. According to their investigation, the solution appears to be sensitive to: (1) choice of the initial trip matrix (2) choice of link flow-delay functions (3) choice of computing correction trip matrix. In particular, the choice of the initial trip matrix is important as the method tends to estimate a final trip matrix which is similar to the starting one. This property led to devoting a substantial attention to constructing a reasonable starting trip matrix. For this purpose, a special gravity type trip distribution model was developed to generate a target trip matrix with desirable attributes. Another sensitive factor to be specified is the selection of link flow-delay functions. Nguyen (1977) proved that the problem solution will replicate observed link flows as long as the link cost functions satisfy two simple criteria: (1) They must be strictly increasing functions of link flows, and (2) they must take the value of observed link cost at the observed link flow. As the best link cost function, Turnquist and Gur used a piece wise linear form, since this function provided superior empirical performance in terms of both the speed of convergence and the quality of the final solutions. Finally, Turnquist and Gur tested several heuristic methods for estimating the correction

trip matrix in (step 5) of the algorithm A4.6. They found that the best approach was to make

$$T_{ij}^* = \begin{cases} T_{ij}^{(n)} + 2(\bar{C}_{ij} - C_{ij}^{(n)}) / (C_{ij}^{(n)} - C_{ij}^{(0)}) & \text{if } \bar{C}_{ij} > C_{ij}^{(n)} \\ 0 & \text{if } \bar{C}_{ij} \leq C_{ij}^{(n)} \end{cases} \quad (4.124)$$

where $C_{ij}^{(0)}$ is the travel cost for i-j pair on the free -flow conditions.

The feasibility of the proposed estimation method was tested in Hudson County, New Jersey. The network had 58 zones and 369 links. In assigning the initial trip matrix to the network by using the equilibrium assignment, a RMS error of 42.7 per cent between the observed flows and the modelled ones was found. The estimated trip matrix, when assigned, showed a RMS error of about 13.5 per cent in link flows.

Finally, as a further improvement to Nguyen's second approach, LeBlanc and Farhangian (1982) developed a more efficient solution method to solve the same estimation problem posed by Gur et al (1979). The problem was solved in two stages. First, Nguyen's second problem P4.18 is solved and the value Z_i^* is found for the objective function $Z_1(\underline{V}, \underline{T})$. Then, in order to choose a trip table to be closest to the target trip matrix the following auxiliary problem was set up:

P4.21

$$\text{Min } \sum_{i,j} (T_{ij} - t_{ij})^2 \quad (4.125)$$

s.t

$$\sum_{r \in P_{ij}} \delta_{ar} h_r = V_a, \quad i=1, \dots, N, j=1, \dots, N. \quad (4.126)$$

$$\sum_{r \in P_{ij}} h_r = T_{ij} \quad (4.127)$$

$$\sum_a \int_0^{V_a} c_a(x) dx - \sum_{i,j} \bar{C}_{ij} T_{ij} \leq Z_i^* \quad (4.128)$$

$$T_{ij}, h_r \geq 0 \quad (4.129)$$

LeBlanc and Farhangian solved P4.21 by taking the partial Lagrangian with respect to the single constraint (4.128). The resulting

Lagrangian problem is:

P4.22

$$\text{Min}_{\underline{T}, \lambda} h(\lambda) = \sum_{i,j} (T_{ij} - t_{ij})^2 + \lambda \left(\sum_a \int_0^{V_a} c_a(x) dx - \sum_{i,j} \tilde{C}_{ij} T_{ij} - Z_1^* \right) \quad (4.130)$$

s.t

$$\sum_{r \in P_{ij}} \delta_{ar} h_r = V_a, \quad \text{for all } i\text{-}j \text{ pairs} \quad (4.131)$$

$$\sum_{r \in P_{ij}} h_r = T_{ij} \quad (4.132)$$

$$T_{ij}, h_r \geq 0 \quad (4.129)$$

Evans's iterative algorithm (Evans, 1976) was used to solve the problem P4.22 for different values of λ . The proposed solution method was tested with a simple network with 24 zones and 76 one-way links on which a known trip matrix had been loaded using equilibrium link assignment and the resulting link flows were used as the observed flows. Two different starting trip matrices were used, resulting in different optimal trip matrices. Both of these optimal trip matrices reproduced very closely the observed link flows. However, both trip matrices were significantly different from the trip matrix originally used to produce the observed flows. Furthermore, the solution method was not considered to be suitable for large networks because of the extensive computing requirement.

4.4.3 Fisk and Boyce's method

Fisk and Boyce (1983) suggested an approach of estimating a trip matrix by calibrating the combined distribution and assignment model using traffic counts. As described in Section 2.3.2, the conventional transport planning process deals with four sequential stages: trip generation, trip distribution, modal split and trip assignment. The process has to be iterated to find a consistent solution in congested networks because of the interdependence between the first three stages and the trip assignment stage. The combined distribution and assignment model treats the second and fourth stages simultaneously thus avoiding the sequential approach for these two stages. The Combined Distribution and Assignment (CDA) problem has been studied by many researchers (for

example, Florian, Nguyen and Ferland, 1975; Evans, 1976 and Erlander, 1977). A CDA problem is equivalent to the following mathematical programming problem (Fisk and Boyce, 1983):

P4.23

$$\text{Min}_{(V_j, T)} \beta \sum_a \int_0^{V_a} c_a(x) dx + \sum_{i,j} T_{ij} \log T_{ij} \quad (4.133)$$

$$\text{s. t.} \quad \sum_i T_{ij} = D_j \quad j=1, \dots, N. \quad (4.134)$$

$$\sum_j T_{ij} = O_i \quad i=1, \dots, N. \quad (4.135)$$

$$\sum_{r \in P_{ij}} h_r = T_{ij} \quad (4.136)$$

$$\sum_{r \in P_{ij}} \delta_{ar} h_r = V_a \quad i=1, \dots, N, j=1, \dots, N. \quad (4.137)$$

$$h_r, T_{ij} \geq 0 \quad (4.138)$$

where β is a parameter to be calibrated.

A number of the solution methods are available for the problem **P4.23**. For example, see Florian et al (1975) and Evans (1976). For a specified value of the parameter β , the optimal solutions of the problem **P4.23** are a doubly constrained gravity model

$$T_{ij} = A_i O_i B_j D_j \exp(-\beta C_{ij}) \quad (4.139)$$

where C_{ij} is the interzonal travel cost on the shortest route, and A_i and B_j are the balancing factors related to the trip end constraints,

and a set of equilibrium link flows consistent with the interzonal travel costs $\{C_{ij}\}$ used in the gravity model.

In order to use the CDA model to estimate a trip matrix, the value of the cost perception parameter β has to be determined a priori. According to Erlander et al (1979), the problem **P4.23** can be reformulated as the following CDA problem:

P4.24

$$\text{Min}_{(\underline{V}, \underline{T})} \sum_{i,j} T_{ij} \log_e T_{ij} \quad (4.140)$$

s. t .

$$\sum_a \int_0^{V_a} c_a(x) dx = C \quad (4.141)$$

$$\sum_i T_{ij} = D_j, \quad j=1, \dots, N \quad (4.142)$$

$$\sum_j T_{ij} = O_i, \quad i=1, \dots, N \quad (4.143)$$

$$\sum_{r \in P_{ij}} h_r = T_{ij} \quad (4.144)$$

$$\sum_{r \in P_{ij}} \delta_{ar} h_r = V_a, \quad i=1, \dots, N, j=1, \dots, N. \quad (4.145)$$

$$h_r, T_{ij} \geq 0 \quad (4.146)$$

It was shown that the problem **P4.24** is equivalent to the problem **P4.23** since the optimal conditions to both problems are same. Furthermore, if $\{V_a^*(\beta)\}$ is the equilibrium flow solution to **P4.24** for a given β value, then, as shown in Erlander et al (1979), the integral

$$\sum_a \int_0^{V_a^*(\beta)} c_a(x) dx = C(\beta) \quad (4.147)$$

is monotone function, implying that β is uniquely determined from **P4.24**. The remaining task is how to estimate the cost $C(\beta)$ using any extra information available. In the past, a sampled trip data was used to calculate the value of $C(\beta)$. Fisk and Boyce (1983) suggested a method of estimating the value of $C(\beta)$ from traffic counts. If observed link flows $\{V_a\}$ are available for all links, $C(\beta)$ can be estimated directly from

$$C(\beta) = \sum_a \int_0^{V_a} c_a(x) dx \quad (4.148)$$

More realistically, only a sample of link counts may be available. Since the sample is typically not random, an unbiased procedure for

estimating the value of $C(\beta)$ is to stratify the sample into K groups of the links classified by link type. Let

$$C_k = \frac{1}{n_k} \sum_{a \in L_k} \int_0^{\bar{v}_a} c_a(x) dx \quad (4.149)$$

be the mean value of the cost function for the group k where L_k is the set of links in group k and n_k is the number of links in L_k . Then, $C(\beta)$ may be estimated as

$$C(\beta) = \sum_{k=1}^K Q_k C_k \quad (4.150)$$

where Q_k is the number of links in group k .

In a recent paper, Fisk (1989) emphasized that this approach requires little computing effort and also is flexible enough to deal with errors in the observed link flows implying the existence of feasible solutions. In particular, this approach is attractive compared to other methods such as those of Willumsen and Nguyen, in the sense that the future link flows can be predicted.

However, despite some advantages advocated by them, Fisk and Boyce's method is just an extension of a doubly constrained gravity model whose applications may not be suitable for urban transport studies. As described in Section 4.3.2.1, the doubly constrained gravity model has the extra total travel cost constraint and an uniform perception to travel costs is assumed. In this respect, their approach is different from entropy maximizing approaches in which no uniform perception to travel costs is assumed. Furthermore, as described above, their approach uses observed flows in a more aggregate way than other methods implying that the estimated trip matrix may not use the information contained in traffic counts fully.

CHAPTER 5. SIMULTANEOUS ESTIMATION OF TRIP MATRICES

5.1 Introduction

As reviewed in Chapter 4, the problem of estimating a trip matrix from traffic counts when the route choice proportions are fixed over traffic demands is well researched. However, when congestion effects in networks play an important role in the traffic assignment, the route choice proportions are not normally fixed and the estimation process becomes more complicated because of the equilibrium conditions. Three methods - Willumsen's method (1981a), Nguyen's method (1977), and Fisk and Boyce's method (1983) - have been developed to tackle this problem. However, as reviewed in Section 4.4, none of these three methods solves the problem satisfactorily. In particular, the sequential method of the ME2 model developed by Willumsen (1981a) appears to be attractive because of its advantages such as the simple data requirement and the low computing cost. On the other hand, the sequential method is only a heuristic, as it solves the equilibrium assignment and the entropy maximization problem alternately. The sequential solution method might fail either to converge or to estimate optimal solutions.

Among other equilibrium based approaches to the estimation problem, Nguyen's method has the form of a traffic assignment problem with elastic demand and it uses a set of the interzonal travel costs as the input data which may be obtained from traffic counts. Fisk and Boyce's method is an extension of a doubly constrained gravity model whose applications may not be suitable for urban transport studies. These two models are distinguished from the ME2 model in the sense that they require the different level of the detail in the input data.

In this chapter, a new formulation and solution method is proposed to estimate a trip matrix from traffic counts under equilibrium traffic conditions, as an alternative to the sequential method of the ME2 model. The new formulation is to maximize entropy values whilst the link flows modelled under equilibrium assignment of the estimated trip matrix reproduce the observed ones. As an objective function in the new

formulation, two different entropy functions, $S_0(\underline{T}, \underline{t})$ and $S_1(\underline{T}, \underline{t})$, are derived and compared. The proposed solution method which solves the equilibrium assignment problem and the matrix estimation problem simultaneously requires a considerable computing demand. A heuristic method is developed which uses a linear approximation fitted by regression to the equilibrium link flows. Extrapolation and perturbation methods have also been used to speed up the solution process. An important feature of the new formulation is that it requires the same input data and can use the same entropy objective function $S_1(\underline{T}, \underline{t})$ as the ME2 model. This enables us to carry out some theoretical and empirical analysis of the differences between the solutions estimated by different methods: the sequential ME2 method and the simultaneous method.

In spite of the use of extrapolation and perturbation methods, the simultaneous method still has a considerable computational requirement for large networks. Because of this, it may be impractical in large networks. An improved sequential method which uses a penalty function method is proposed. This method estimates an optimum solution by approaching the feasible region progressively, while fixed route choice proportions \underline{P} are used.

This chapter is organized as follows. Section 5.2 describes a new formulation of the matrix estimation method. The new formulation will be compared with that of the ME2 model. Section 5.3 describes a simultaneous solution procedure proposed to solve the new formulation. A constraint approximation method and the sequential unconstrained optimization technique are described there in detail. In order to reduce the heavy computational demand of the proposed solution method, some heuristic methods such as an extrapolation method and a perturbation method are described. Finally, Section 5.4 describes an improved sequential estimation method.

5.2 Formulation

The simultaneous formulation of estimating trip matrices from traffic counts adopts the same underlying idea as the ME2 model but with two distinct modifications to the ME2 model. These are: (1) the new formulation can use either of the entropy functions $S_0(\underline{T}, t)$ and $S_1(\underline{T}, \underline{t})$ as an objective function in the formulation of the problem. (2) the new formulation sets assigned link flows equal to observed link flows instead of using the assignment proportions of the estimated trip matrices. In the following sub-sections, each of these modifications is described in detail.

5.2.1 Entropy objective function

Entropy is most commonly known in the physics and engineering fields in connection with the second law of thermodynamics - the entropy law. This states that the entropy, or amount of disorder, in any closed conservative thermodynamic system tends to increase. A fundamental step in using entropy in information theory was provided by Shannon (1948). Shannon showed that entropy, which measures the amount of disorder in a thermodynamic system, also measures the amount of entropy or uncertainty in a probabilistic sense. Shannon discovered that the amount of entropy in any discrete probability distribution is proportional to

$$- \sum_{i=1}^N p_i \log_e p_i \quad (5.1)$$

where p_i is the probability of event i and $\sum_{i=1}^N p_i = 1$.

Shannon's entropy function (5.1) has several useful properties. For any fixed N , Equation (5.1) is a continuous and symmetric concave function with respect to all its arguments. It has first and second derivatives and has a global maximum when all the probabilities are equal.

The next key step in opening up new applications for the Shannon measure of entropy was taken by Jaynes (1957). Jaynes suggested that the Shannon entropy measure could be used in a reverse sense to generate or

infer a probability distribution which would have maximum entropy. Since entropy is a measure of uncertainty, a maximum entropy distribution must have maximum uncertainty, must be maximally non-committal and must therefore contain minimum bias. If we have any partial information about some random process, we should not arbitrarily choose some probability distribution to fit but we should choose that probability distribution which maximizes the Shannon entropy measure subject to the partial information we have. The probability distribution which results from this constrained maximization process will then be one which introduces minimum bias into the probability estimation process. Jaynes's work is often referred to as the Maximum Entropy Principle.

As already reviewed in Section 4.3.2, Wilson (1967; 1970) adapted the maximum entropy principle to transport and regional planning problems. Wilson derived a doubly constrained gravity model by maximizing an entropy measure subject to trip end and total cost constraints. Following the derivations by Wilson (1970), and Batty and March (1976), we here derive measures of entropy associated with a trip matrix $\underline{T}=\{T_{ij}\}$ and optionally with a prior trip matrix $\underline{t}=\{t_{ij}\}$.

The number of micro-states associated with a trip matrix $\{T_{ij}\}$ is given by

$$W(\underline{T}) = \frac{T_{..}!}{\prod_{i,j} T_{ij}!} \quad (5.2)$$

where $T_{..} = \sum_{i,j} T_{ij}$.

If a prior trip matrix \underline{t} is available, (5.2) becomes

$$W(\underline{T}, \underline{t}) = \frac{T_{..}! \prod_{i,j} \left(\frac{t_{ij}}{t_{..}}\right)^{T_{ij}}}{\prod_{i,j} T_{ij}} \quad (5.3)$$

where $t_{..} = \sum_{i,j} t_{ij}$

Taking the logarithm of (5.3),

$$\log_e W(\underline{T}, \underline{t}) = \log_e T_{..}! + \sum_{i,j} T_{ij} (\log_e t_{ij} - \log_e t_{..}) - \sum_{i,j} \log_e T_{ij}! \quad (5.4)$$

Using Stirling's formula: $\log_e T_{ij}! \approx T_{ij}(\log_e T_{ij} - 1)$, we obtain a full entropy function from (5.4),

$$S_0(\underline{T}, \underline{t}) = T_{..}(\log_e T_{..} - 1) - \sum_{i,j} T_{ij} (\log_e (T_{ij}/t_{ij}) - 1 + \log_e t_{..}) \quad (5.5)$$

If we assume that the total demand $T_{..}$ is fixed, we have the simplified entropy function

$$S_1(\underline{T}, \underline{t}) = - \sum_{i,j} T_{ij} (\log_e T_{ij}/t_{ij} - 1) \quad (5.6)$$

$$= S_0(\underline{T}, \underline{t}) + K \quad (5.7)$$

where K is a constant.

If we ignore prior information \underline{t} , that is, set $t_{ij}=1$ for all i and j , $S_0(\underline{T}, \underline{t})$ and $S_1(\underline{T}, \underline{t})$ become

$$S_0(\underline{T}) = T_{..}(\log_e T_{..} - 1) - \sum_{i,j} T_{ij} (\log_e T_{ij} - 1) \quad (5.8)$$

and

$$S_1(\underline{T}) = - \sum_{i,j} T_{ij} (\log_e T_{ij} - 1) \quad (5.9)$$

As reviewed in Section 4.3.2.1, Bell (1983) also derived a full entropy function $S_0(\underline{T}, \underline{t})$ based on the assumption that the trips of T_{ij} are multinomially distributed. Bell derived $S_0(\underline{T}, \underline{t})$ by maximizing the joint probability of observing the trip matrix \underline{T} given the prior trip matrix \underline{t} . As shown in Equations (5.6) and (5.7), the entropy function $S_1(\underline{T}, \underline{t})$ can be given as a simplified form of $S_0(\underline{T}, \underline{t})$ by assuming that the total demand $T_{..}$ is constant. In particular, the ME2 model uses $S_1(\underline{T}, \underline{t})$ as an entropy objective function. In deriving the ME2 model, Willumsen (1981a) assumed that the total demand $T_{..}$ is fixed. However, in practice the total demand $T_{..}$ is unknown and so it should be treated as a variable than a constant. In addition to that, it can be shown that the full entropy function $S_0(\underline{T}, \underline{t})$ differs from the simplified one

$S_1(\underline{T}, t)$ with respect to the following properties:

(1) The value of $S_0(\underline{T}, t)$ is invariant to the application of uniform scaling to the prior trip matrix. That is, $S_0(\underline{T}, t) = S_0(\underline{T}, \lambda t)$, for all $\lambda > 0$, where λ is a scaling factor. It was revealed in Section 4.3.2.1 that the use of $S_1(\underline{T}, t)$ in the ME2 model fails to estimate unbiased solutions, especially when the significant change in traffic demand is made since the estimation of the prior trip matrix. On the other hand, the use of $S_0(\underline{T}, t)$ allows control over the total demand of the estimated trip matrix and accordingly such bias found in the solutions produced by the ME2 model could be avoided.

(2) $S_1(\underline{T}, t)$ is the sum of a number of strictly convex functions and so it is strictly convex (Willumsen, 1981a). On the other hand, it can be proven that $S_0(\underline{T}, t)$ is convex but not strictly (Bell, 1983). Furthermore, the entropy function $S_0(\underline{T}, t)$ has the maximum value $S_0(\underline{T}, t) = 0$ at the stationary points $\underline{T} = \lambda t$, where $\lambda > 0$.

5.2.2 Link flow constraints

The estimated trip matrix can only be constrained from observed link flows through the traffic assignment process. In the case of capacity restrained equilibrium assignment, the route choice proportions \underline{P} of the trips are not constant when the trip matrix varies. Furthermore, the route choice proportions \underline{P} are not uniquely determined by the equilibrium assignment process. For these reasons, we propose an alternative approach to the method of using the route choice proportions. Instead of using the route choice proportions, the new approach sets assigned equilibrium link flows equal to observed ones. That is

$$V_a^*(\underline{T}) = \hat{V}_a, \quad a \in I \quad (5.10)$$

where $V_a^*(\underline{T})$ is assigned equilibrium flows on link a ,

\hat{V}_a is observed link flows on link a ,

I is the set of links for which observations are made.

5.2.3 A new formulation

Following the entropy maximization approach taken by Wilson (1967) and Willumsen (1981a), the problem formulated is to maximize the entropy measures subject to assigned link flows reproducing observed ones when the estimated trip matrix is assigned to the network. The proposed formulation (Oh, 1989a) is

P5.1

$$\text{Max}_{\underline{T}} S_0(\underline{T}, \underline{t}) \text{ or } S_1(\underline{T}, \underline{t}) \quad (5.11)$$

s.t.

$$V_a^*(\underline{T}) = \bar{V}_a, \quad a \in I \quad (5.12)$$

The problem P5.1 is a single optimization problem containing both trip matrix estimation and, in the constraints, equilibrium assignment. It is expected to have better convergence properties than the ME2 method, especially in congested networks. Furthermore, as equilibrium link flows are determined uniquely by equilibrium assignment, it does not suffer from the ill determination of the route choice proportions \underline{P} . This implies that the solutions obtained from the new formulation might yield higher entropy values than the ME2 model. Finally, the problem P5.1 can use either the full entropy function $S_0(\underline{T}, \underline{t})$ which does not assume that the total demand $T_{..}$ is fixed or the simplified one $S_1(\underline{T}, \underline{t})$. Because of this, the problem P5.1 produces unbiased solutions in the same counter-example used by Maher (1987).

5.3 A simultaneous solution method

The problem P5.1 is an optimization problem with a convex objective function and non-linear equality constraints. Furthermore, the equilibrium link flows $\{V_a^*(\underline{T})\}$ in the constraints of the problem P5.1 cannot be expressed in a closed functional form, but rather are found by solving equilibrium assignment problems. Any solution method which requires analytical expressions for gradients of constraints cannot be used. This leads to an application of the sequential unconstrained optimization methods (Luenberger, 1984). These methods approximate the

solutions of the original constrained optimization problems by solving a sequence of unconstrained ones in which the objective function has been modified to penalize to some extent any violation of the constraints of the original problem. Its main advantage is that the constraints need not be dealt separately and any methods for solving unconstrained optimization problems can be applied. In the next sub-section, sequential unconstrained optimization methods are introduced in detail and their application to Problem P5.1 is described.

5.3.1 Use of sequential unconstrained optimization methods

Sequential unconstrained optimization methods, which are also known as the penalty and barrier methods, are procedures for approximating constrained optimization problems by solving a sequence of unconstrained ones. A sequence of unconstrained problems is generated from the original constrained problem by sequentially increasing the magnitude of the penalty parameter by which the penalty function of the constraints is multiplied. The approximation is accomplished in the case of the penalty method by adding to the objective function a penalty term that prescribes a high cost for violation of the constraints. In the case of the barrier method, a similar effect is achieved by adding a term that favors points interior to the feasible region over those near the boundary. In general, the barrier method is used when the objective function is not defined outside the feasible region. In the case of the problem P5.1, the penalty function method is more suitable.

Problem P5.1 is transformed into a sequence of unconstrained sub-problems. This transformation is done by defining an appropriate auxiliary function - the penalty function - and a series of the penalty parameter values. The penalty function $G(\underline{T}, \underline{V})$ is chosen amongst those which satisfy the following properties (Luenberger, 1984, pp 365-395):

- (1) $G(\underline{T}, \underline{V})$ is continuous
- (2) $G(\underline{T}, \underline{V}) \geq 0$ for all \underline{T}
- (3) $G(\underline{T}, \underline{V}) = 0$ iff $\underline{T} \in \underline{S}$, where \underline{S} is the feasible solution set.

This study uses a quadratic function as the penalty function, while other forms such as a likelihood function are also possible. The quadratic penalty function is simple and easy to use. The resulting transformed problem to Problem P5.1 (Oh, 1989a) is:

P5.2

$$\underset{\underline{T}}{\text{Max}} S_0(\underline{T}, \underline{t}) + \mu_n G(\underline{T}, \underline{V}) \quad (n=0,1,2,\dots) \quad (5.13)$$

where $G(\underline{T}, \underline{V}) = \sum_{a \in I} (V_a^*(\underline{T}) - \underline{V}_a)^2$ is the penalty function representing the gap between the modelled flows and the observed ones, and the penalty parameter μ_n ($n=0,1,2,\dots$) is negative and decreasing in n .

The penalty function $G(\underline{T}, \underline{V})$ in Problem P5.2 satisfies the properties required to the penalty function. Unlike the Lagrangian method, the penalty function method uses fixed values for μ_n within each subproblem. As the penalty parameter μ_n decreases sequentially, the solution points will approach the feasible region and also converge to a solution which is also optimal for the original problem P5.1.

The values of the penalty function parameter μ_n in Problem P5.2 are specified externally and there are a number of possibilities to set them. In general, a sequence of the parameter values to Problem P5.2 is generated from the following relationship

$$\mu_{n+1} = \mu_n c, \quad \text{where } c > 1 \text{ and } n=0,1,2,\dots \quad (5.14)$$

The solutions and the computing times incurred to solve the problems is sensitive to the following factors: (1) the choice of the initial parameter value μ_0 , (2) the choice of the value of the multiplier c and (3) the number of the subproblems to be solved. The initial penalty parameter value μ_0 is normally chosen to be such as $|\mu_0| < \epsilon$ where ϵ is small enough and the value of the multiplier c is set to be such as $c > 1$. The number of the subproblems required to reach the final solution depends on the initial penalty parameter value μ_0 and the value of the multiplier c . In particular, when the problems solved are ill-conditioned or have multiple optima, better solutions might be

achieved by starting with a small initial penalty parameter value such as $\mu_0 = \pm 10^{-3}$ and then to be increased in magnitude gradually by using the value of the multiplier c such as $c=2$. According to Fiacco and McCormick (1968, pp 188-191), the extrapolation method introduced later in Section 5.3.4 can be a powerful computational tool to approximate the final solution using the sequence of the solutions already obtained. This method helps to accelerate the solution process, especially when started with small initial penalty parameter values.

Problem P5.2 is one of unconstrained non-convex maximization. It can be solved by the sequential application of any uni-dimensional line search method. Line search methods such as the golden section search method are procedures for solving one-dimensional or higher dimensional optimization problems by executing a sequence of successive line searches (Luenberger, 1984, pp 199-200). However, their application to Problem P5.2 leads to a considerable computational requirement, especially in view of the number of equilibrium sub-problems that are required to be solved. Furthermore, the non-linear nature of equilibrium link flow constraints results in non-convexity and this can lead to ill-conditioning of the problem. At least, there could exist multiple local optimum solutions for Problem P5.2.

Some of these difficulties can be alleviated by adopting a suitable functional form as an approximation to actual equilibrium link flows. The requirements for such an approximation are that it should be convenient to fit, reasonably accurate, and simple in form. Once fitted, approximation of this kind can be evaluated at minimal computational cost and can also be used to provide gradients for use in solution methods. In the next section, a method to approximate equilibrium link flows is presented.

At this stage, it is necessary to explain the distinction between the ME2 method and the simultaneous method in using the word 'sequential'. The ME2 method solves the two sub-problems of matrix estimation problem and equilibrium assignment sequentially. On the other hand, the simultaneous method solves two sub-problems simultaneously by executing a sequence of multi-dimensional optimisation problems, each of

which respects the interaction between the matrix estimation and the assignment problems

5.3.2 Linear approximation to equilibrium link flows

In order to reduce the computational burden and to overcome the non-convexity of the penalty function $G(\underline{T}, \underline{V})$, a heuristic method has been developed to approximate the equilibrium link flows by estimating linear functions of each T_{ij} (Oh, 1989b; 1989c).

Let $\underline{V}^*(\underline{T}) = \underline{P}(\underline{T})\underline{T}$ be the equilibrium link flows generated by the trip matrix \underline{T} . Let $\underline{\delta T}$ be a perturbation from the trip matrix \underline{T} . In the following, the trip matrix \underline{T} is assumed to be perturbed for each i - j pair in turn and then the element δT_{ij} in the perturbation matrix $\underline{\delta T}$ is such that

$$|\delta T_{ij}| = \delta_{oi}\delta_{dj}|\underline{\delta T}| \quad (5.15)$$

where $\delta_{oi}\delta_{dj} = 1$, if $i=o$ and $j=d$, and $\delta_{oi}\delta_{dj} = 0$, otherwise.

Then, applying Taylor's formula, we have the following polynomial expression

$$\underline{V}^*(\underline{T}+\underline{\delta T}) = \underline{P}(\underline{T})\underline{T} + (\underline{T} \frac{\partial \underline{P}(\underline{T})}{\partial \underline{T}} + \underline{P}(\underline{T}))\underline{\delta T} + \underline{R}_2 \quad (5.16)$$

where \underline{R}_2 is of second order in $|\underline{\delta T}|$.

If $|\underline{\delta T}|$ is small enough, the residual \underline{R}_2 can be ignored. Then, Equation (5.16) becomes

$$\underline{V}^*(\underline{T}+\underline{\delta T}) \cong \underline{P}(\underline{T})\underline{T} + (\underline{T} \frac{\partial \underline{P}(\underline{T})}{\partial \underline{T}} + \underline{P}(\underline{T}))\underline{\delta T} \quad (5.17)$$

so,

$$\begin{aligned} \underline{V}_a^*(\underline{T}+\underline{\delta T}) &= \alpha_{aij} + \underline{\beta}_a \underline{\delta T} & a \in I \\ &= \alpha_{aij} + \beta_{aij} \delta T_{ij} & a \in I \end{aligned} \quad (5.18)$$

where $\{\alpha_{aij}\}$ and $\{\beta_{aij}\}$ are parameters to be estimated for each (i,j) pair.

Equation (5.18) is a first-order Taylor series approximation to the true relationship Equation (5.16). The simultaneous solution method using this approximate relationship is contrasted with the sequential ME2 method in a respect that the approximation relationship includes the term $\partial P(\underline{T})/\partial \underline{T}$ in Equation (5.17). Because of this, the simultaneous solution method retains the dependence between the trip matrix and equilibrium link flows.

In Equation (5.17), the coefficients, $\{\alpha_{aij}\}$ and $\{\beta_{aij}\}$, are estimated for each of observed links $a \in I$ by using least squares estimation method over a set of particular values of $\underline{T} + \delta \underline{T}$ and equilibrium link flows using $\underline{T} + \delta \underline{T}$. This is done anew for each (i,j) pair. The number of points for regression is to be specified. In principle, the goodness of fit between equilibrium link flows and approximated ones is subject to the combination of two major errors: the truncation error due to the lack of convergence in the equilibrium assignment, and the mis-specification error due to the assumption of the simple linear model. These two major errors are systematic rather than random in the sense that they can be controlled by taking some measures. For example, the truncation error can be reduced as the convergence criterion of the assignment algorithm is made more stringent but at a high cost in computing time. On the other hand, the mis-specification error can be reduced by increasing the number of points used in the regression but again at a high cost in computing time. Under the circumstance of the limited computing resources, the goodness of fit may be improved by allocating the computing resources optimally between the two measures: thus the choice lies between performing more assignment iterations at fewer points and performing fewer assignment iterations at more points. However, there is no apparent optimum strategy for this allocation. For a small range of perturbations, the mis-specification error is considered to be small and thus two regression points will be enough. On the other hand, as the range of perturbations increases, more than two points will be useful to improve the goodness of fit.

5.3.3 A solution algorithm

Replacing the constraints of Problem P5.2 with Equation (5.18), we obtain the following modified problem:

P5.3

$$\text{Max}_{\underline{\delta T}} S_0(\underline{T} + \underline{\delta T}, t) \quad (5.19)$$

s.t.

$$\alpha_{aij} + \beta_a \underline{\delta T} = V_a, \quad a \in I \quad (5.20)$$

Transforming Problem P5.3 into a sequence of unconstrained problems by using the penalty function method, we have:

P5.4

$$\text{Max}_{\underline{\delta T}} \text{So}(\underline{T}+\underline{\delta T}, \underline{t}) + \mu_n G(\underline{T}+\underline{\delta T}, \underline{V}) \quad (5.21)$$

$$\text{where } G(\underline{T}+\underline{\delta T}, \underline{V}) = \sum_{a \in I} (\alpha_{aij} + \beta_a \underline{\delta T} - \underline{V}_a)^2$$

Problem P5.4 is one of a sequence of unconstrained optimization problems and is convex for each uni-dimensional line search. Problem P5.4 which is a multi-dimensional problem can be solved by performing a sequence of uni-dimensional line searches or uni-dimensional gradient searches. Differentiating the objective function $O(\underline{T}+\underline{\delta T}) = \text{So}(\underline{T}+\underline{\delta T}, \underline{t}) + \mu_n G(\underline{T}+\underline{\delta T}, \underline{V})$ of Problem P5.4 with respect to δT_{ij} , we obtain the following first-order necessary conditions:

$$\frac{\partial O(\underline{T}+\underline{\delta T})}{\partial \delta T_{ij}} = \log_e(T_{..} + \delta T_{ij}) - \log_e(T_{ij} + \delta T_{ij}) + \log_e t_{ij} - \log_e t_{..} + 2\mu_n \sum_{a \in I} \beta_{aij} (\alpha_{aij} + \beta_a \underline{\delta T} - \underline{V}_a) = 0 \quad (5.22)$$

Equation (5.22) is monotonic but non-linear, though the solutions $T_{ij} + \delta T_{ij}^*$ to Equation (5.22) can be obtained efficiently by a root finding method.

The remainder of this section describes the algorithm used to solve the problem P5.4. The main process is based on the sequential unconstrained optimization structure. As shown in Figure 5.1, it searches for the solution to a constrained problem by finding a sequence of solutions to unconstrained subproblems, as the values of the penalty parameter μ_n changes. The algorithm includes the following major steps:

A5.1

(step 1) Set the initial penalty parameter value μ_0 and initialize $\underline{T} = \underline{t}$.

(step 2) Set a feasible search interval $[X_L, X_U]$ for each T_{ij} .

(step 3) Set $i=1$ and $j=1$.

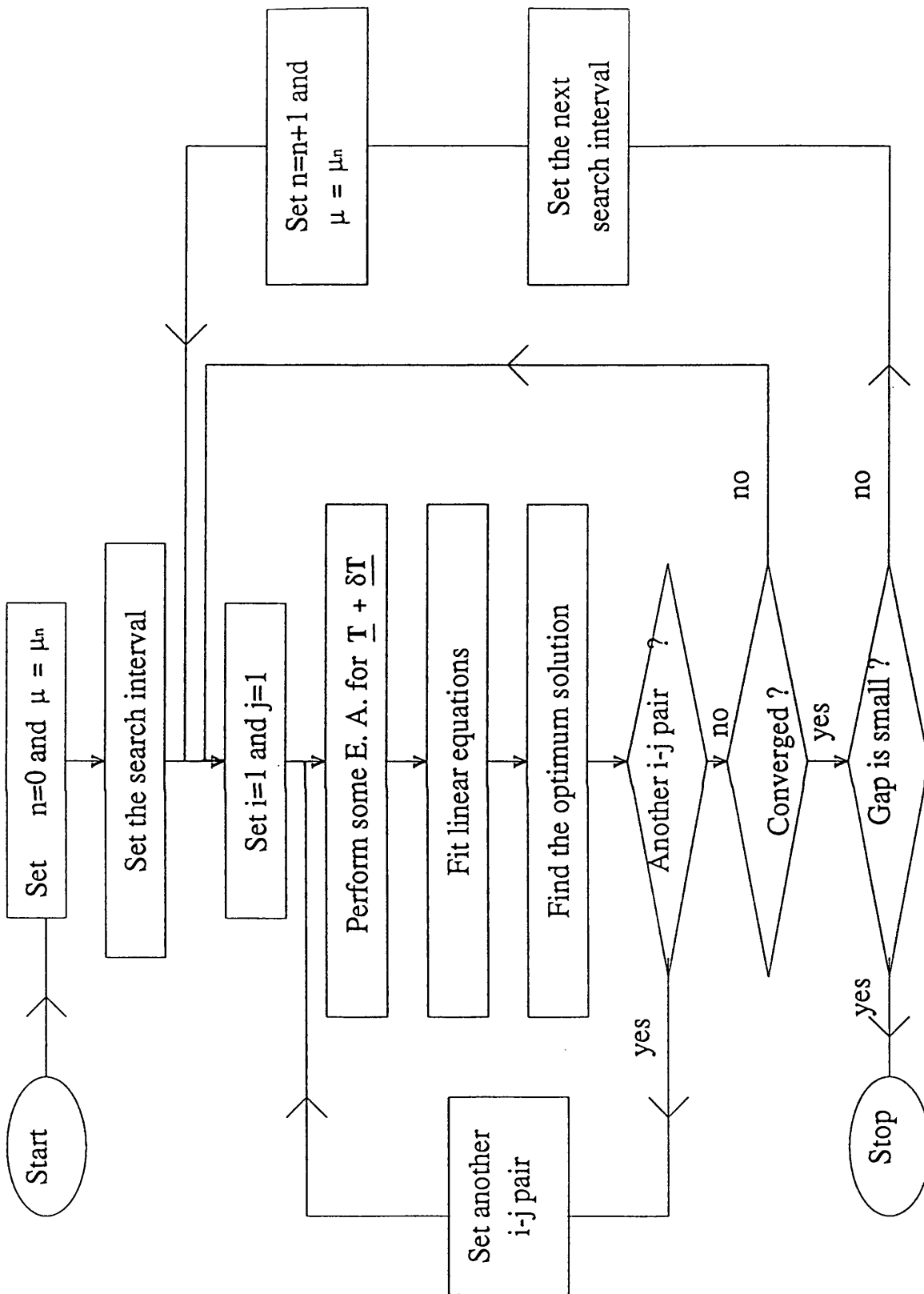


Figure 5.1 Flow chart of the simultaneous solution method (A5.1)

(step 4) Select five values of δT_{ij} from the specified search interval. Perform equilibrium assignment at each of five values using the perturbation method described in Section 5.3.5 and regress the resulting equilibrium link flows over five corresponding values of δT_{ij} to fit the approximate linear relationships $V_a^*(T+\delta T) = \alpha_{a ij} + \beta_a \delta T$, $a \in I$.

(step 5) Solve Equation (5.22) for the optimal value δT_{ij}^* .

(step 6) If there is another O-D pair to be considered, set another O-D pair and return to (step 4). Otherwise, check for convergence. If convergence is achieved, proceed to (step 7). Otherwise, return to (step 3).

(step 7) Check gap, the value of the penalty function $G(\underline{T}, \underline{V})$. If the gap value is sufficiently close to zero, stop. Otherwise, set the next search interval using the extrapolation method described in Section 5.3.4 and increase the magnitude of the penalty parameter and return to (step 3).

In (step 4) of the algorithm A5.1, a different number of values of δT_{ij} other than five values could be specified for regression. Also, in (step 5), it is possible that the solutions are on the boundary of the search interval and hence better solutions lie outside the search interval. If this occurs, the search interval is extended to include the better solutions.

The algorithm A5.1 uses a sequential approach in solving the multi-dimensional problem. This performs sequentially a number of uni-dimensional searches, while each uni-dimensional search solves entropy maximization and equilibrium assignment simultaneously. Therefore, the proposed algorithm is simultaneous in the sense that the dependence of the equilibrium link flows and the trip demand is respected during the optimization. Although the sequential solution procedure in solving the multi-dimensional problem will ultimately converge, the successful progress of the proposed solution procedure will depend on the goodness of fit between equilibrium link flows from

the assignment and approximate link flows from the regressed linear relationship. In particular, the goodness of fit between two flows much depends on the degree of convergence to equilibrium link flows in the assignment.

As discussed in Section 3.3.4, traffic counts are not normally error-free and they are unlikely to be consistent with the traffic assignment model. It was revealed that because of this there might be no feasible solutions to the matrix estimation problem. The penalty function method approximates the final solutions by progressively increasing the magnitude of the penalty parameter value. This allows the simultaneous solution method proposed in this study to be more flexible in dealing with inconsistent link flows than the existing sequential ME2 method: the process can be terminated at any stage if no further improvement in fit is forthcoming.

Another important feature of the simultaneous solution method is that it is modular and so it can accommodate other objective functions or penalty functions without fundamental modification. This will be discussed further in Section 7.3 of Chapter 7.

5.3.4 Use of an extrapolation method

For the sake of reducing the computing complexity, two heuristic methods are adopted in the solution method. The first one is the use of an extrapolation method developed by Fiacco and McCormick (1968, pp 188-191). Fiacco and McCormick suggested that as a powerful computational tool, the extrapolation method can be used to accelerate convergence when solving constrained optimization problems using sequential unconstrained method. This method is based on the existence of a unique trajectory of constrained local maxima. It uses information on such a trajectory to estimate the solution to the next subproblem and the final solution. In particular, the extrapolation method is used to establish the starting point and the initial search interval for the local maximum of the next sub-problem, as mentioned in the simultaneous solution algorithm A5.1.

Following Fiacco and McCormick (1968, pp 188-191), suppose that $\underline{X}(\mu_i)$ is the solution of the i th sub-problem using the penalty parameter μ_i . Also, suppose that $\underline{X}(\mu_i)$ can be approximated using the polynomial formula:

$$\underline{X}(\mu_i) = \sum_{j=0}^{p-1} \underline{a}_j(\mu_i)^j \quad (5.23)$$

where \underline{a}_j is the vector of j th coefficient.

Then, the limiting solution as $\mu_i \rightarrow -\infty$ can be approximated by \underline{a}_0 in Equation (5.23). However, it is not possible to calculate the value of \underline{a}_0 from (5.23). As a practical method, Fiacco and McCormick developed a simple recursive scheme for computing a series of estimates of the limiting solution. This scheme is based on a particular structure of the penalty parameter, that is, $\mu_i = \mu_0 c^i$ where $c > 1$.

After p sub-problems have been solved, approximation to $\underline{X}(\mu_i)$ can be made using polynomials of order j ($1 \leq j \leq p-1$). Let \underline{X}_{ij} ($1 \leq i \leq p$, $0 \leq j \leq i-1$) signify the estimate of $\underline{X}(-\infty)$ using solutions to sub-problems ($p-(i-1), p-(i-2), \dots, p$) and a polynomial approximation of order j .

The best approximation to $\underline{X}(-\infty)$ is then given by

$$\underline{X}(-\infty) \approx \underline{X}_{p,p-1} \quad (5.24)$$

$$= \underline{a}_0 \quad (5.25)$$

In the case where $\mu_i = \mu_0 c^i$ ($c > 1$), the values \underline{X}_{ij} can be calculated without estimating \underline{a}_k ($0 \leq k \leq j$) by use of the recursive formula:

$$\underline{X}_{ij} = \frac{c^j \underline{X}_{i,j-1} - \underline{X}_{i-1,j-1}}{c^j - 1} \quad (2 \leq i \leq p, 1 \leq j \leq i-1) \quad (5.26)$$

$$\underline{X}_{i,0} = \underline{X}(\mu_i) \quad (1 \leq i \leq p) \quad (5.27)$$

The solution to sub-problem $p+1$, $\underline{X}(\mu_{p+1})$, can be estimated by $\underline{X}_{p+1,0}$ using the approximation that $\underline{X}_{p+1,p-1} \cong \underline{X}_{p,p-1}$. Again, using $\mu_{p+1} = \mu_1 c^p$, this is done by the recursive formula:

$$\underline{X}_{p+1,j-1} = \frac{(c^j-1) \underline{X}_{p+1,j} + \underline{X}_{p,j+1}}{c^j} \quad (5.28)$$

When sub-problem $p+1$ has been solved to give an accurate value for $\underline{X}(\mu_{p+1})$, the value of p can be incremented and the formula (5.26) used again.

The estimate $\underline{X}_{p+1,0}$ and the error between the previous maximum $\underline{X}_{p,0}$ and the estimate $\underline{X}_{p,0}$, $|\underline{X}_{p,0}-\underline{X}_{p,0}|$, can be used to establish the next search interval using a value such as $\underline{X}_{p+1,0} \pm |\underline{X}_{p,0}-\underline{X}_{p,0}|$.

5.3.5 Use of the perturbation method

The second heuristic is the use of a perturbation method in solving the equilibrium assignment problem, as used in the simultaneous solution algorithm **A5.1**. First, the method calculates feasible link flows, $\underline{V}(\underline{T}+\underline{\delta T})$, for a perturbation $\underline{\delta T}$ to the trip matrix \underline{T} from the previous equilibrium link flows $\underline{V}^*(\underline{T})$, rather than performing an equilibrium assignment from the beginning. Starting with perturbed link flows, some iterations of the usual Frank-Wolfe equilibrium assignment procedure are performed to yield new equilibrium link flows $\underline{V}^*(\underline{T}+\underline{\delta T})$.

There are many ways depending on the rules of calculating the feasible link flows from the previous equilibrium link flows. In this study, two ways are considered and the first one was used for the tests described later in Chapter 6. The first one is to use the previous route choice proportions \underline{P} . That is:

$$\underline{V}(\underline{T}+\underline{\delta T}) = \underline{V}^*(\underline{T}) + \underline{P} \underline{\delta T} \quad (5.29)$$

where $\underline{V}(\underline{T}+\underline{\delta T})$ is the vector of perturbed link flows.

This method is simple to understand, but it requires the retention of the whole array of the route choice proportions \underline{P} which makes this perturbation method impractical for large networks.

An alternative method is to use the minimum cost paths and the maximum cost paths for assigning the perturbed trip matrix. The proposed solution method **A5.1** considers in turn each element of the trip matrix \underline{T} . Let δT_{ij} be an element of $\underline{\delta T}$. Then it follows that

$$|\delta T_{ij}| = \delta_{oi}\delta_{dj}|\underline{\delta T}| \quad (5.30)$$

where $\delta_{oi}\delta_{dj}=1$, for $o=i$ and $d=j$, otherwise, $\delta_{oi}\delta_{dj}=0$.

We separate two cases, positive δT_{ij} and negative δT_{ij} , and apply different rules for calculating $\underline{V}(\underline{T}+\underline{\delta T})$ from $\underline{V}^*(\underline{T})$.

Case I: Positive δT_{ij}

A5.2

(step 1) Find the minimum cost paths, p_{ij}^* , using the previous link costs, $\underline{c}(\underline{V}^*(\underline{T}))$ between zone i and zone j .

(step 2) Add δT_{ij} to link flows $\underline{V}(\underline{T})$ only on the minimum paths and thus produce $\underline{V}(\underline{T}+\underline{\delta T})$, i.e.

$$\underline{V}(\underline{T}+\underline{\delta T}) = \underline{V}^*(\underline{T}) + \underline{V}(\underline{\delta T}) \quad (5.31)$$

$$\text{where } \underline{V}(\underline{\delta T}) = \{V_a(\underline{\delta T})\} \text{ and } V_a(\underline{\delta T}) = |\underline{\delta T}| \delta_{ap_{ij}^*}$$

$$\text{where } \delta_{ap_{ij}^*} = 1 \text{ if link } a \text{ lies on path } p_{ij}^*$$

$$= 0 \text{ otherwise}$$

(step 3) Perform some iterations of the equilibrium assignment to get $\underline{V}^*(\underline{T}+\underline{\delta T})$.

Case II: Negative δT_{ij}

A5.3

(step 0) Set $\Delta T = |\underline{\delta T}|$.

(step 1) Find the maximum cost paths, p_{ab}^+ , using only links of positive link flows, such as $\underline{V}^*(\underline{T}) > 0$. Let V_{\min} be the minimum link flow on the maximum cost paths. The maximum cost path is the most

expensive path amongst the feasible ones satisfying the following conditions:

- (1) No link in the path is used more than once,
- (2) zone centroids are not used in the path, except as the origin and destination.

(step 2) If $\Delta T > V_{\min}$, deduct V_{\min} from all the links on the maximum cost paths. Set $\Delta T = \Delta T - V_{\min}$ and return to (step 1). Otherwise $\Delta T \leq V_{\min}$, deduct ΔT from all the links on the maximum cost path.

(step 3) Perform some iterations of the equilibrium assignment starting with the feasible link flows $\underline{V}(\underline{T} + \delta T)$ to get the equilibrium link flows $\underline{V}^*(\underline{T} + \delta T)$.

5.3.6 Interface with an equilibrium assignment program

The proposed solution method performs many equilibrium assignments in order to calculate the equilibrium link flows for the estimated trip matrices. This can be achieved efficiently by establishing an interface between the solution methods and an existing equilibrium assignment program. In this study, the solution method has been interfaced with the equilibrium assignment program of the SATURN suite (Van Vliet, 1987).

SATURN is a computer model for the analysis and evaluation of traffic management schemes. SATURN uses two models, the assignment model and the simulation model, in order to achieve realistic assignments in networks. The assignment model is used to obtain an estimate of flows in links and the simulation model is used to estimate capacity, queues and delays at junctions. As shown in Figure 4.2 in Section 4.4.1, several iterations of the assignment-simulation cycle could be performed to produce a self-consistent set of flows and costs. However, in this study, only the assignment model has been considered and delays are represented by link-based user defined speed-flow relationships. As a traffic assignment model, SATURN provides most features of standard assignment packages such as generalized cost, all-or-nothing, Wardrop's equilibrium, stochastic assignment, etc. SATURN is also linked with the

ME2 model for estimating trip matrices from traffic counts. In particular, the ME2 model in SATURN can be used for comparison.

5.3.7 Computational demand of the simultaneous solution method

The heavy computational demand of the simultaneous solution method appears to be the main drawback. The simultaneous solution method requires many equilibrium assignments to be performed. The number of equilibrium assignments required depends on the network size and the level of congestion of the network. For example, in a network with 30 zones, the number of equilibrium assignments required for performing one iteration of the optimisation for all T_{ij} cells of the trip matrix would be about 4350 if one uses 5 points for the linear regression for each T_{ij} . The amount of computing time required to solve each equilibrium assignment depends on the level of congestion in the network and the network size. If the network is more congested, more Frank-Wolfe iterations will be needed for convergence to the equilibrium link flows. In order to reduce the computing burden, the perturbation and extrapolation methods are applied.

In spite of the use of the perturbation method and the extrapolation method, the simultaneous method still has an uncertainty about the practicality in computing for problems with large networks. However, computing capability is improving rapidly and this advance in computing technology will permit the simultaneous method to be more practical in the near future.

5.4 An improved sequential solution method

The simultaneous solution method proposed to solve two sub-problems of matrix estimation and equilibrium assignment simultaneously appears to be impractical because of the heavy computing demand in large networks. This led to the development of an alternative estimation method which can be practical in terms of computing demand and also can deal with congestion effects reasonably well.

The sequential solution method used in the ME2 model estimates a trip matrix from traffic counts under traffic equilibrium conditions by solving the two sub-problems of equilibrium assignment and entropy maximization alternately. The main feature of that method is to use the route choice proportions \underline{P} which are in general not fixed over variation in trip demand in congested networks. The sequential method intends ultimately to find mutually consistent equilibrium route choice proportions and trip matrix given the traffic counts. The prior trip matrix is used to generate initial route choice proportions which are then used to estimate an improved trip matrix. In general, the convergence of this sequential solution method depends on the degree of the coupling between the route choice proportions and the trip matrix estimation. One would expect that the convergence of the sequential method (ME2) might be hampered by attempting to satisfy the feasibility condition from the start. Any large change in the trip matrix will lead to large changes in the route choice proportions and again the large changes in the route choice proportions leads to the large changes in the estimated trip matrix. This process could continue until reasonable convergence is achieved. Better convergence might be achieved by imposing the feasibility condition progressively rather than from the start, while the route choice proportions are still being used. As a way to impose the feasibility conditions progressively, the sequential unconstrained optimisation method already used for the simultaneous solution method can be used to solve the problem. The main difference between the improved sequential method and the simultaneous one is that the improved sequential method uses the route choice proportions, whilst the simultaneous method uses approximate link flows in the constraints.

The improved sequential solution method uses a similar formulation to the ME2 model but a different solution method. This does use the route choice proportions to relate the estimated trip matrix to observed link flows.

P5.5

$$\text{Max}_{\underline{T}} S_0(\underline{T}, \underline{t}) \text{ or } S_1(\underline{T}, \underline{t}) \quad (5.32)$$

$$\text{s.t. } \underline{P}\underline{T} = \underline{V} \quad (5.33)$$

Problem **P5.5** includes convex constraints with fixed route choice proportions \underline{P} . Instead of using Lagrangian and multi-proportional methods, the sequential unconstrained optimisation method is applied to solve the problem **P5.5**. As described in Section 5.3.1, the sequential unconstrained optimisation method approximates the optimum solution by approaching the feasible region progressively. Problem **P5.5** is transformed into a sequence of unconstrained problems.

P5.6

$$\text{Max } S_o(\underline{T}, \underline{t}) + \mu_n G_1(\underline{T}, \underline{V}) \quad (n=0,1,2,\dots) \quad (5.34)$$

where $G_1(\underline{T}, \underline{V}) = \sum_a \sum_{i,j} (P_{ij}^a T_{ij} - \bar{V}_a)^2$ is a penalty function and the penalty parameter μ_n ($n=0,1,2,\dots$) is negative and decreasing in n . Also, the penalty function $G_1(\underline{T}, \underline{V})$ in the problem **P5.6** satisfies the properties required to the penalty function.

Each of unconstrained problems in **P5.6** can be solved by using the first order necessary optimum conditions. Differentiating the objective function $O_1(\underline{T}) = S_o(\underline{T}, \underline{t}) + \mu_n G_1(\underline{T}, \underline{V})$ of Problem **P5.6** with respect to T_{ij} , we obtain:

$$\frac{\partial O_1(\underline{T})}{\partial T_{ij}} = \log T_{e..} - \log T_{eij} + \log t_{ij} - \log t_{e..} + 2\mu_n \sum_{a \in \{1, \dots, d\}} P_{ij}^a (\sum_{o,d} P_{od}^a T_{od} - \bar{V}_a) = 0 \quad (5.35)$$

Equation (5.35) is non-linear and the solutions to Equation (5.35) can be obtained efficiently by using a root finding method.

The solution algorithm proposed for the simultaneous method can be used after making only minor modifications. The modified algorithm whose flow chart is shown in Figure 5.2 includes the following major steps:

A5.4

(step 1) Set the initial penalty parameter value μ_0 and initialize $\underline{T} = \underline{t}$.

(step 2) Set a feasible search interval $[X_L, X_U]$ for each T_{ij} .

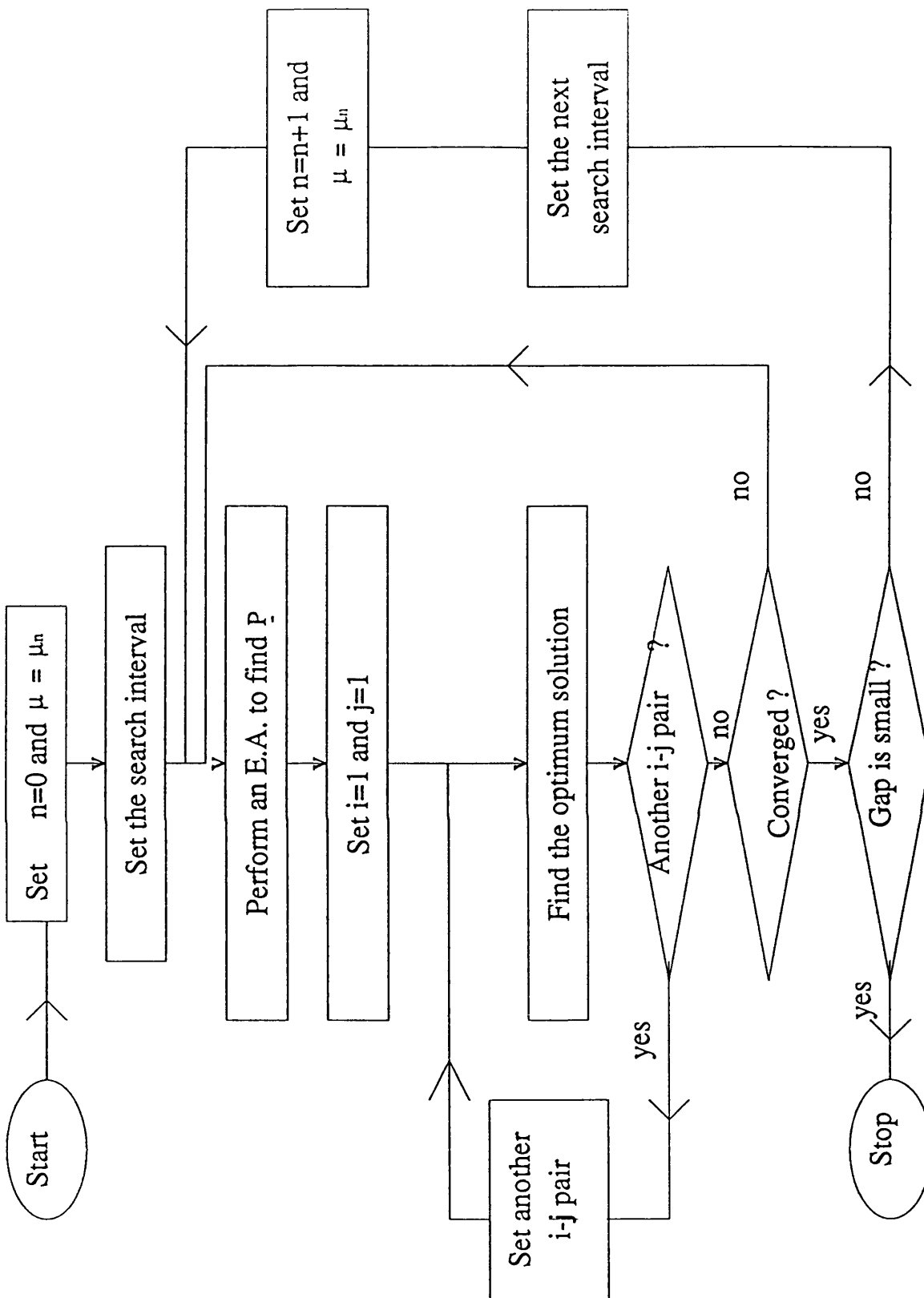


Figure 5.2 Flow chart of the improved sequential solution method (A5.4)

(step 3) Perform the equilibrium assignment for the trip matrix \underline{T} using the perturbation method and retain the route choice proportions \underline{P} .

(step 4) Solve Equation (5.35) for the optimal values $\{T_{ij}^*\}$ iteratively.

(step 5) Check for convergence. If convergence has been achieved, proceed to (step 6). Otherwise, return to (step 3) to update the route choice proportions \underline{P} using the improved trip matrix.

(step 6) Check gap, the value of the penalty function $G_1(\underline{T}, \underline{V})$. If the gap value is sufficiently close to zero, stop. Otherwise, set the next search interval using the extrapolation method and decrease the penalty parameter value and return to (step 3).

The modified algorithm A5.4 greatly reduces the computing time by performing a single equilibrium assignment for all T_{ij} values rather than performing several equilibrium assignments for each T_{ij} . The modified solution method is sequential as it uses the fixed route choice proportions. An important feature of this improved sequential method is that it is practical in terms of computing time, although more equilibrium assignments are required compared with the ME2 method.

As described in Section 5.3.3, the improved sequential solution method also has the same features as does the simultaneous one: (1) it can accommodate inconsistent link flows without any prior modification, and (2) it is modular and so it can be applied with other objective functions or penalty functions.

CHAPTER 6. TESTS OF THE ESTIMATION METHODS

6.1 Introduction

The sequential method (Hall, Van Vliet and Willumsen, 1980) to estimate trip matrices from traffic counts under traffic equilibrium conditions is only a heuristic, as it solves the two-subproblems of equilibrium assignment and entropy maximization alternately. It is noted (Willumsen, 1982; Fisk, 1988) that the sequential method might fail either to converge or to estimate optimal solutions. As an alternative to the sequential method, a new formulation and solution method has been proposed in Chapter 5. The proposed solution method solves the problems of equilibrium assignment and entropy maximization simultaneously. However, it is noted that the heavy computational demand is a major difficulty facing the simultaneous method. For the sake of reducing the computational burden, an improved sequential method which takes advantages of some features of the sequential unconstrained optimization method (the penalty function method) was also proposed in Chapter 5.

The main objective of this chapter is to test the proposed estimation methods. In order to do this, three test cases are designed and tested and the results are reported. Another important objective of this chapter is to provide a good understanding of the solutions by the sequential method. This is done by comparing the solutions by different estimation methods: the sequential method, the improved sequential method and the simultaneous method.

This chapter is organized as follows. Section 6.2 reports on tests using a simple network. This is designed to investigate some theoretical aspects of the solutions by the sequential method. Section 6.3 reports on tests using a more complicated example network taken from Nguyen (1977). The test is designed to investigate various interesting topics on the performance of the estimation methods. These include the congestion effects and the inconsistency in traffic counts. Finally, Section 6.4 reports on tests using the real network and data collected in the town of Ripon in North Yorkshire. The tests in Section 6.4 are

designed to investigate how well various estimation methods perform in the real network.

6.2 Tests with a simple network

6.2.1 Introduction

A simple network was prepared to have the following features: (1) route choice in the network should be sufficiently sensitive to traffic demand and (2) the network should be simple enough to admit an explicit relationship to calculate the equilibrium link flows. As shown in Figure 6.1, the simple network was designed to have two origin zones, one destination zone and two links. Origin zone 1 is connected to both links but origin zone 2 is connected only to link 2. Therefore, trips between zone 1 and zone 3, T_{13} can use either of two routes. However, trips between zone 2 and zone 3, T_{23} , can use only one route.

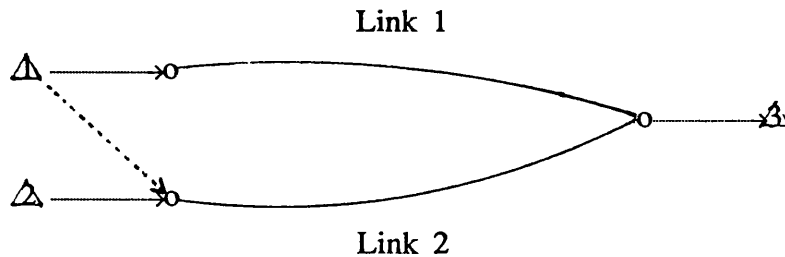


Figure 6.1 The simple example network

Capacity restraint in the network is achieved by separable link-based speed-flow functions whereby the travel time on each link is assumed to be a function of the flows only on that link. As shown in Figure 6.2, the travel time on link 1 was designed to be less than that on link 2 until an equilibrium link flow, V^* , is reached. The speed-flow curve in link 2 was designed to be constant with respect to traffic demand. Thus, trips T_{13} use only link 1 until V^* is reached and then all further flows use link 2. In other words, the maximum allowed link flows in link 1 is V^* .

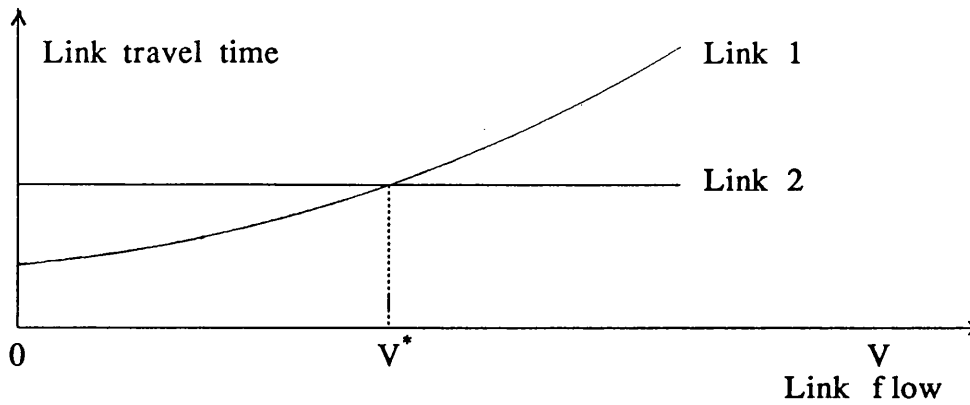


Figure 6.2 The speed-flow curve for the simple network

Using the assignment program of the SATURN Suite (Van Vliet, 1987), for the specific network input data given, the equilibrium link flow, $V^* \cong 877$ (veh/h), was obtained.

In this example, an explicit form of the equilibrium link flow conditions could be used in the calculation. This is useful, as the equilibrium link flows for the estimated trip matrix can be obtained without performing the Frank-Wolfe algorithm, although this is not possible in more complicated networks. These explicit equilibrium conditions are expressed as follows:

$$V_1^* = \text{Min}(877, T_{13}) \quad (6.1)$$

$$V_2^* = T_{23} + \text{Max}(0, T_{13} - 877) \quad (6.2)$$

where V_1^* is the equilibrium flow on link 1,

V_2^* is the equilibrium flow on link 2.

Another feature of this example is that the route choice proportions P are determined uniquely for any given trip matrix by the equilibrium traffic assignment. It follows that the sequential method using this simple network works on unique route choice proportions, although this cannot be done in most cases.

Trip matrices containing two non-zero values are estimated from one observed link flow on link 2. Thus, the estimation problem is underspecified. If observed link flows are available for both links, the problem is not fully underspecified, as the total number of the estimated trip matrix is known. Two different set of observed link flows have been used. The lower flow, $\bar{V}_2=500$ (veh/h) was chosen to be lower than $V^*=877$, and $\bar{V}_2=1100$ (veh/h) was chosen to be higher than $V^*=877$.

Trip matrices are estimated by three solution methods: the sequential method, the simultaneous method using $S_1(\underline{T}, \underline{t})$ as the objective function, the simultaneous method using $S_0(\underline{T}, \underline{t})$. For the sequential method, the computer programs - SATASS and SATME2 - in the SATURN programs (Van Vliet, 1987) were used to estimate trip matrices. The main procedure to estimate trip matrices from traffic counts in the SATURN programs has been described in Section 4.4.1. In particular, the specification of an initial trip matrix used to generate the initial route choice proportions could be important, as the use of the different initial matrices leads to the different estimated ones. Normally, the prior trip matrix is used as the initial trip matrix, if it is available. If it is not specified, a uniform prior matrix possibly calculated from traffic counts is specified.

The simultaneous solutions have been produced analytically by using the Lagrange method to the following optimization problem.

P6.1

$$\text{Max } S_1(\underline{T}, \underline{t}) \text{ or } S_0(\underline{T}, \underline{t}) \tag{6.3}$$

s.t.

$$T_{23} + \text{Max}(0, T_{13}-877) = \bar{V}_2 \tag{6.4}$$

where $S_1(\underline{T}, \underline{t}) = -\sum_{i,j} T_{ij}(\log_e (T_{ij}/t_{ij})-1)$

$$S_0(\underline{T}, \underline{t}) = T_{..}(\log_e T_{..}-1) - \sum_{i,j} T_{ij}(\log_e (T_{ij}/t_{ij})-1 + \log_e t_{.j})$$

Case 1: The simultaneous solutions (SIM-S1) using $S_1(\underline{T}, \underline{t})$.

If $T_{13} \leq 877$, **P6.1** becomes

P6.2

$$\text{Max}_{\underline{T}} S_1(\underline{T}, \underline{t}) \quad (6.5)$$

$$\text{s.t. } T_{23} = \bar{V}_2 \quad (6.6)$$

By using the Lagrange method, **P6.2** becomes

$$L(\underline{T}, \lambda) = - \sum_{i,j} T_{ij} (\log_e(T_{ij}/t_{ij}) - 1) - \lambda (T_{23} - \bar{V}_2) \quad (6.7)$$

where λ is a Lagrange multiplier.

The necessary conditions for the stationary points are

$$\frac{\partial L}{\partial T_{13}} = -\log_e(T_{13}/t_{13}) = 0 \quad (6.8)$$

$$\frac{\partial L}{\partial T_{23}} = -\log_e(T_{23}/t_{23}) - \lambda = 0 \quad (6.9)$$

$$\frac{\partial L}{\partial \lambda} = -(T_{23} - \bar{V}_2) = 0 \quad (6.10)$$

By solving (6.8), (6.9) and (6.10) simultaneously, the stationary points are

$$T_{13} = t_{13}, \quad T_{23} = \bar{V}_2 \quad (6.11)$$

If $T_{13} > 877$, **P6.1** becomes

P6.3

$$\text{Max}_{\underline{T}} S_1(\underline{T}, \underline{t}) \quad (6.12)$$

$$\text{s.t. } T_{23} + T_{13} - 877 = \bar{V}_2 \quad (6.13)$$

By the same process as used for **P6.2**, we obtain the following stationary points for the problem **P6.3**.

$$T_{13} = \frac{t_{13}}{(t_{13}+t_{23})}(877+\bar{V}_2), \quad T_{23} = \frac{t_{23}}{(t_{13}+t_{23})}(877+\bar{V}_2) \quad (6.14)$$

If $t_{13} > 877$, the solution (6.11) becomes infeasible, and if $T_{13} = \frac{t_{13}}{(t_{13}+t_{23})}(877+\bar{V}_2) < 877$, the solution (6.14) becomes infeasible. If (6.11) and (6.14) are infeasible, we have a non-stationary optimum point at the extreme point, $(877, \bar{V}_2)$.

Case 2: The simultaneous solutions (SIM-So) using $So(\underline{T}, \underline{t})$.

If $T_{13} \leq 877$, **P6.1** becomes

P6.4

$$\text{Max } So(\underline{T}, \underline{t}) \quad (6.15)$$

$$\text{s.t. } T_{23} = \bar{V}_2 \quad (6.16)$$

By the same process as used for **Case 1**, we obtain the following stationary points for the problem **P6.4**.

$$T_{13} = \frac{t_{13}}{t_{23}}\bar{V}_2, \quad T_{23} = \bar{V}_2 \quad (6.17)$$

If $T_{13} > 877$, **P6.1** becomes

P6.5

$$\text{Max } So(\underline{T}, \underline{t}) \quad (6.18)$$

$$\text{s.t. } T_{23} + T_{13} - 877 = \bar{V}_2 \quad (6.19)$$

By the same process as used for **Case 1**, we obtain the following stationary points for the problem **P6.5**.

$$T_{13} = \frac{t_{13}}{(t_{13}+t_{23})}(877+\bar{V}_2), \quad T_{23} = \frac{t_{23}}{(t_{13}+t_{23})}(877+\bar{V}_2) \quad (6.20)$$

The solution (6.17) becomes infeasible if $\frac{t_{13}}{t_{23}}\bar{V}_2 > 877$, and the solution (6.20) becomes infeasible if $\frac{t_{13}}{(t_{13}+t_{23})}(877+\bar{V}_2) < 877$. However, at least one of these points is feasible.

6.2.2 Results and comments

The results are presented in Table 6.1. Table 6.1 shows detailed estimates produced from using different prior trip matrices, different observed link flows and different solution methods: the sequential method (ME2), the simultaneous method using $S_1(\underline{T}, t)$ (SIM-S₁) and the simultaneous method using $S_0(\underline{T}, t)$ (SIM-S₀). The results in Table 6.1 can be summarized as follows. These results are discussed in terms of optimality, stationarity and convergence of the solution method. In Table 6.1, the annotation 'optimum' means that the solution achieved is optimum, 'stationary' means that a stationary optimum solution is achieved, 'converged' means that the solution achieved is non-optimum, 'not stationary' means that the solution achieved is non-stationary optimum and 'not converged' means that the solution method fails to converge, etc.

(1) The sequential method ME2 converges to the values $(T_{13}, T_{23}) = (t_{13}, \bar{V}_2)$, whenever $t_{13} \leq 877$. In this case the estimate (t_{13}, \bar{V}_2) is optimal and stationary with respect to the objective function $S_1(\underline{T}, t)$. However, if $t_{13} > 877$, the sequential solution method either does not converge or converges to a feasible solution, which is not optimal in all cases.

(2) The simultaneous method using the objective function $S_1(\underline{T}, t)$ has the stationary local optimum solution (t_{13}, \bar{V}_2) , whenever $t_{13} \leq 877$. However, if $t_{13} > 877$, it has the stationary local optimum solutions, $[(877 + \bar{V}_2)t_{13}/(t_{13} + t_{23}), (877 + \bar{V}_2)t_{23}/(t_{13} + t_{23})]$, and the global optimum solution $(877, \bar{V}_2)$ which is not stationary.

(3) The simultaneous method using the objective function $S_0(\underline{T}, t)$ has the two stationary local optimum solutions $(\bar{V}_2 t_{13}/t_{23}, \bar{V}_2)$, and $[(877 + \bar{V}_2)t_{13}/(t_{13} + t_{23}), (877 + \bar{V}_2)t_{23}/(t_{13} + t_{23})]$ for any values of t_{13} . In particular, the use of a uniform prior trip matrices leads to estimates with an equal distribution, that is (\bar{V}_2, \bar{V}_2) for $\bar{V}_2 \leq 877$ and $[(877 + \bar{V}_2)/2, (877 + \bar{V}_2)/2]$ for $\bar{V}_2 > 877$.

Table 6.1 Results of the tests using the simple example network

(a) $\bar{V}_2=500$ (veh/h) and uniform-prior trip matrices

Solution Methods	Trip Matrix			Entropy		Notes
	T ₁₃	T ₂₃	T _{..}	S _o	S ₁	
Prior	400	400	800	0	800	-
ME 2	400	500	900	-6	788	optimum
SIM - S ₁	400	500	900	-6	788	stationary
SIM - S _o	500	500	1000	0	777	stationary
Prior	1000	1000	2000	0	2000	-
ME 2	966	458	1424	-93	1815	converged
SIM - S ₁	877	500	1377	-52	1838	not stationary
SIM - S _o	500	500	1000	0	1693	stationary
Prior	1100	1100	2200	0	2200	-
ME 2	-	-	-	-	-	not converged
SIM - S ₁	877	500	1377	-52	1970	not stationary
SIM - S _o	500	500	1000	0	1788	stationary

(b) $\bar{V}_2=500$ (veh/h) and non-uniform prior trip matrices

Solution Methods	Trip Matrix			Entropy		Notes
	T ₁₃	T ₂₃	T _{..}	S _o	S ₁	
Prior	400	1000	1400	0	1400	-
ME 2	400	500	900	-51	1247	optimum
SIM - S ₁	400	500	900	-51	1247	stationary
SIM - S _o	200	500	700	0	1185	stationary
Prior	800	400	1200	0	1200	-
ME 2	800	500	1300	-8	1188	optimum
SIM - S ₁	800	500	1300	-8	1188	stationary
SIM - S _o	918	459	1377	0	1188	stationary
Prior	1000	400	1400	0	1400	-
ME 2	994	383	1377	0	1400	optimum
SIM - S ₁	994	383	1377	0	1400	stationary
SIM - S _o	994	383	1377	0	1400	stationary
Prior	1000	1200	2200	0	2200	-
ME 2	-	-	-	-	-	not converged
SIM - S ₁	877	500	1377	-92	1930	stationary
SIM - S _o	417	500	917	0	1719	stationary
Prior	1200	900	2100	0	2100	-
ME 2	-	-	-	-	-	not converged
SIM - S ₁	877	500	1377	-12	1946	not stationary
SIM - S _o	667	500	1167	0	1853	stationary

Table 6.1 (cont.)

(c) $V_2=1100$ (veh/h) and uniform prior trip matrices

Solution Methods	Trip Matrix			Entropy		Notes
	T ₁₃	T ₂₃	T _{..}	S _o	S ₁	
Prior	700	700	1400	0	1400	-
ME2	700	1100	1800	-45	1303	optimum
SIM - S ₁	700	1100	1800	-45	1303	stationary
SIM - S _o	988	988	1976	0	1295	stationary
Prior	1000	1000	2000	0	2000	-
ME2	998	981	1979	-1	1999	converged
SIM - S ₁	988	988	1976	0	2000	stationary
SIM - S _o	988	988	1976	0	2000	stationary
Prior	1200	1200	2400	0	2400	-
ME2	1118	861	1979	-17	2344	converged
SIM - S ₁	988	988	1976	0	2360	stationary
SIM - S _o	988	988	1976	0	2360	stationary

(d) $V_2=1100$ (veh/h) and non-uniform prior trip matrices

Solution Methods	Trip Matrix			Entropy		Notes
	T ₁₃	T ₂₃	T _{..}	S _o	S ₁	
Prior	400	1000	1400	0	1400	-
ME2	400	1100	1500	-1	1395	optimum
SIM - S ₁	400	1100	1500	-1	1395	stationary
SIM - S _o	440	1100	1540	0	1393	stationary
Prior	800	400	1200	0	1200	-
ME2	800	1100	1900	-240	787	local optimum
SIM - S ₁	1318	659	1977	0	990	stationary
SIM - S _o	1318	659	1977	0	990	stationary
Prior	900	1300	2100	0	1400	-
ME2	897	1082	1979	-8	2181	converged
SIM - S ₁	877	1100	1977	-5	2183	not stationary
SIM - S _o	761	1100	1861	0	2172	stationary
Prior	1000	400	1400	0	1400	-
ME2	1194	783	1977	-55	1239	converged
SIM - S ₁	1412	565	1977	0	1295	stationary
SIM - S _o	1412	565	1977	0	1295	stationary
Prior	1200	1500	2700	0	2700	-
ME2	1086	891	1977	-44	2549	converged
SIM - S ₁	879	1098	1977	0	2593	stationary
SIM - S _o	879	1098	1977	0	2593	stationary

The following comments are made on these results.

(1) It has been found empirically that the sequential method does not always converge, and even if it does converge, it generally yields non-optimum solutions whenever $t_{13} > 877$. This result can be interpreted as follows. The condition $t_{13} > 877$ means that traffic demand of T_{13} uses both routes. The route choice proportions become variable between successive iterations of trip matrix estimation and equilibrium assignment. However, the route choice proportions used are assumed to be fixed by the ME2 process and this is imposed implicitly as an extra constraint during each trip matrix estimation.

(2) In the sequential method, the total demand $T_{..}$ is assumed to be fixed during the trip matrix estimation. It has been found empirically that during the operation of the sequential method, this assumption is not valid.

(3) In this example, the route choice proportions resulting from equilibrium assignment are uniquely specified. If this property does not obtain, the sequential method might well produce poorer results.

(4) Even in this simple example, the problem of estimating the trip matrix does not have a single relative optimum. That is, the problem is not convex. This fact implies that the simultaneous method might identify one of many stationary points.

6.3 Tests with Nguyen's network

6.3.1 Introduction

In the previous section, a simple example network was used to investigate various solutions estimated by the sequential method and the simultaneous one. In that case, it was possible to obtain the equilibrium link flows for the estimated trip matrices explicitly and without performing equilibrium assignment. However, this is not possible in more complicated networks. Instead, the proposed solution methods need to be interfaced with a suitable equilibrium assignment program in order to calculate the equilibrium link flows for the estimated trip matrices. For this purpose, in this study the SATURN assignment program (Van Vliet, 1987) was interfaced with the simultaneous solution method. In particular, the use of the SATURN program is helpful as it also includes a matrix estimation procedure of using the sequential ME2 method.

In this section, a more complicated artificial example network is used to test the simultaneous solution method and the improved sequential solution method and to make comparison with the solutions by the sequential solution method. The use of an artificial network and associated data base in testing the methods has both advantages and disadvantages. As major advantages, it gives the analyst full control over data errors and a variety of different cases can be investigated, and also it is useful to test the computer program. In this study, a small artificial example is useful to explore various aspects of the proposed estimation methods with only limited computing effort. However, the artificial data will not be realistic in all respects and the test using the artificial data can only give a general indication of how the proposed methods will perform in practice.

The example network used in this test, as shown Figure 6.3, is based on the test network used by Nguyen (1977). The network consists of 2 origin zones, 2 destination zones, 9 nodes, 11 one-way links and 8 zone centroids connectors. Only four origin-destination pairs $\{(1,3),(1,4),(2,3),(2,4)\}$ allocate flow to the network: for convenience, the trip matrix is denoted by an ordered set of 4 O-D pairs

corresponding to $(T_{13}, T_{14}, T_{23}, T_{24})$. Although it is a small network, it appears to be suitable for a test network, as it has a number of alternative routes between each of the four O-D pairs when the network becomes congested. The table of the network data used to specify link cost functions in SATURN is given in Appendix 2.

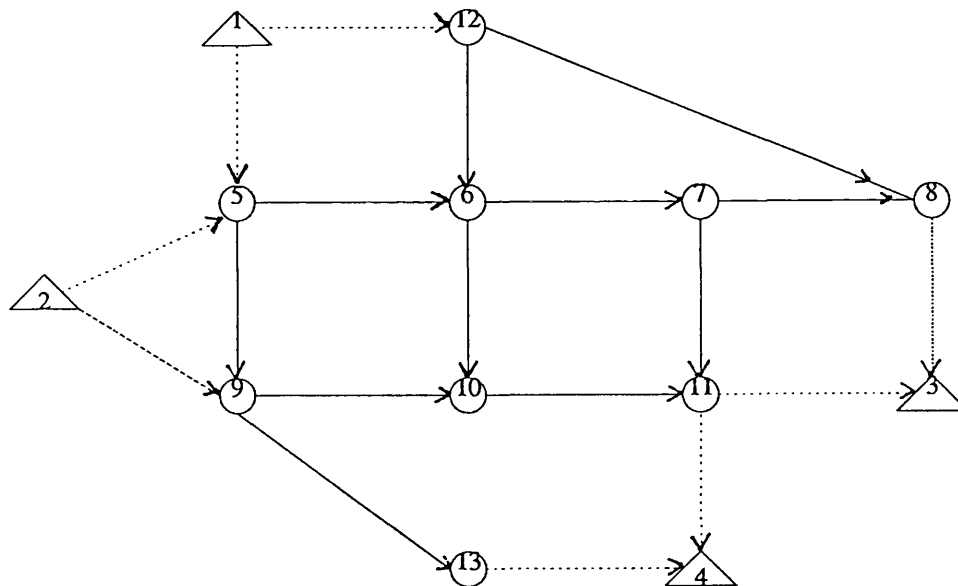


Figure 6.3 Nguyen's (1977) example network

6.3.2 Objectives and design of tests

The main objectives of the tests are to answer the following two questions: (1) How well do the proposed solution methods perform in estimating a known trip matrix from a set of observed link flows? and (2) How different are the trip matrices estimated by the simultaneous method, the improved sequential method and the sequential one?

First of all, some main elements which might affect the results of the tests are considered in order to answer these questions. These are:

(1) **The choice of the solutions methods:** Three solution methods are used to estimate a trip matrix from traffic counts. These are: the sequential ME2 method (ME2), the improved sequential method (ME3) and the simultaneous method (SIM).

(2) **The choice of the entropy objective functions:** Two different entropy functions combined with the prior trip matrices: $S_1(\underline{T}, t)$ and $S_0(\underline{T}, t)$ are used as the objective functions for ME3 and SIM. Thus, ME3- S_1 denotes the improved sequential method using the entropy objective function $S_1(\underline{T}, t)$, ME3- S_0 denotes the improved sequential method using $S_0(\underline{T}, t)$, SIM- S_1 denotes the simultaneous method using $S_1(\underline{T}, t)$ and SIM- S_0 denotes the simultaneous method using $S_0(\underline{T}, t)$. The sequential ME2 method implemented in SATURN (Van Vliet, 1987) can only be used with the objective function $S_1(\underline{T}, t)$.

(3) **The congestion levels in the network:** Various levels of network congestion are considered to investigate the effects of congestion on the performance of the various estimation methods.

(4) **The level of errors in traffic counts:** Various levels of inconsistency in link flows are generated to investigate the effects of errors in traffic counts.

(5) **The number of links observed:** different numbers of observed links are considered to investigate its effects.

(6) **The use of different prior trip matrices.**

(7) **The selection of links observed.**

Starting from these seven elements, the following tests were designed:

- (1) Base test
- (2) Tests with various congestion levels
- (3) Tests with inconsistent link flows
- (4) Tests with determinate matrix total
- (5) Tests with different prior trip matrices
- (6) Tests with different sets of links observed.

Each of these tests is described in detail together with the numerical results in the following sections.

In the following tests, the performance of the estimation methods will be assessed on the basis of three major abilities: **convergence** of the solution method, **optimality** of the solution achieved and **goodness of fit** of the estimated trip matrix to a known one.

(1) **Convergence:** Convergence is the ability to approach a solution as the number of iterations increases. This can be checked by investigating discrepancy (gap) between the modelled link flows and the observed link flows. Discrepancy (gap) is measured by the root mean square error (RMSE) between these two link flows:

$$\text{Gap} = \left[\frac{1}{M} \sum_{a \in I} (V_a^* - V_a)^2 \right]^{1/2} \quad (6.21)$$

(2) **Optimality:** Optimality is the ability to find a feasible solution which maximises (or minimises) the objective criterion. This can be checked by examining the entropy measures: $S_1(\underline{T}, t)$ and $S_0(\underline{T}, t)$.

(3) **Goodness of fit:** Goodness of fit is the ability how closely the estimated trip matrix is to the observed trip matrix. This is measured by the root mean square errors between the two matrices:

$$\text{Fit} = \left[\frac{1}{N(N-1)} \sum_{i, j} (T_{ij} - \hat{T}_{ij})^2 \right]^{1/2} \quad (6.22)$$

In each of the tests, the measures to use for comparison are as follows.

(1) **Estimated trip matrices:** the cell values of the estimated trip matrices for each of five estimation methods are compared and convergence of matrix cell values is examined as the number of iterations or the magnitude of the penalty parameter increases.

(2) **Total demand:** the total demands of the estimated trip matrices for each of the five estimation methods are compared and their convergence is examined as the number of iterations or the magnitude of the penalty parameter increase.

(3) **Entropy:** two measures of entropy calculated from each of two

different entropy functions $S_0(\underline{T},t)$ and $S_1(\underline{T},t)$ are compared.

(4) **Gap:** RMSE values between the modelled link flows and the observed ones are compared.

(5) **Goodness of fit(Fit):** RMSE values between the estimated trip matrices and the observed trip matrices are compared.

6.3.3 Base test

6.3.3.1 Introduction

The performance of the estimation methods might be affected by various elements such as congestion effects, errors in link flows, the number of links observed, the prior trip matrices and selection of links observed. A number of test cases can be designed by various ways of combining these elements. The base test is designed as a case which is considered to be representative of real problems. It estimates a trip matrix from a data set which simulates a realistic situation by including various elements properly. In this sense, the base test is compared with other test cases in which a single particular aspect of the estimation methods is tested. For example, when we investigate the effects of congestion on the performance of the estimation methods, error-free observed link flows are used.

An artificial data set was constructed to test the base case. First, a trip matrix, $(T_{13},T_{14},T_{23},T_{24})=(300,300,300,300)$ was chosen as a known observed trip matrix to generate a certain level of network congestions for the test. The level of congestion on links is measured by a value of V/C , an average ratio between link flows and the link capacity on links observed. In this test, a value of $V/C=0.8$ was chosen. Error-free link flows were produced from the observed trip matrix by equilibrium assignment. Four links - (6,7), (6,10), (9,10), (10,11) - are selected as observed link flows. Because of continuity of flows at node 10, as shown in Figure 6.3, if there were no errors in observed link flows, one of the three counts on links (6,10), (9,10) and (10,11) is redundant. Inconsistency in observed link flows is generated by adding random quantities within the range of $\pm 20\%$ of the calculated

equilibrium link flows using a rectangular distribution. Finally, a trip matrix (350,250,150, 50) which is not a uniform scaling of the observed matrix was chosen as a prior trip matrix.

The performance of the various estimation methods is investigated by comparing the cell values of estimated trip matrices, total demand, entropy objective values, gap values and goodness of fit of estimated trip matrices for each of five estimation methods: ME2, ME3-S₁, SIM-S₁, ME3-S₀ and SIM-S₀.

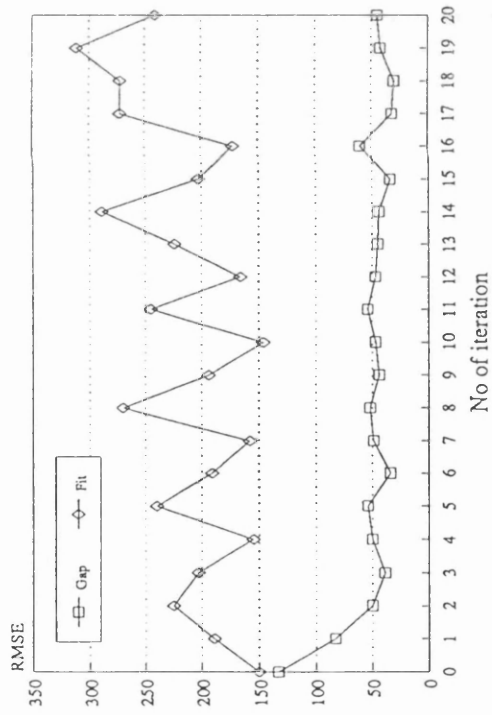
6.3.3.2 Results

It is interesting to see the performance of the sequential estimation method of ME2 as the number of iterations between equilibrium assignment and matrix estimation increases. The results of this test - estimated trip matrices, entropy values, gap values and the goodness of fit of estimated matrices - are presented in Table 6.2. The evolution of performance indicators presented in Table 6.2 are depicted graphically in Figures 6.4a-6.4d. It can be seen from Figure 6.4a that the sequential method reduces gap rapidly during the first three iterations and makes relatively small changes to it thereafter.

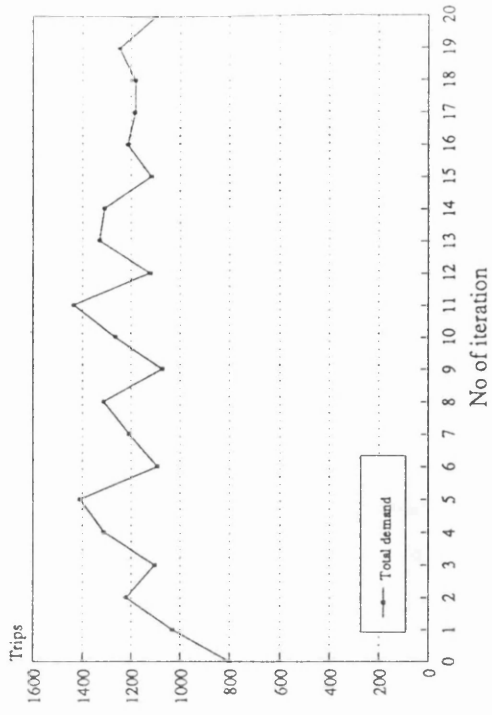
Table 6.2 Results of the base test by ME2

Iteration	Trip Matrix					Entropy		Gap		Fit	
	T13	T14	T23	T24	T..	S0	S1	RMSE	%RMSE	RMSE	%RMSE
Observed	300	300	300	300	1200	-267	446	39	12	0	0
Prior	350	250	150	50	800	0	800	133	42	150	50
1	523	127	329	50	1030	-123	646	83	26	189	63
2	644	164	359	52	1219	-126	580	50	16	225	75
3	450	104	502	47	1104	-270	478	39	12	203	68
4	435	349	454	73	1310	-94	570	50	16	155	52
5	692	240	418	58	1409	-103	509	54	17	240	80
6	470	122	451	50	1093	-207	545	34	11	191	64
7	494	257	386	70	1207	-69	641	49	16	158	53
8	746	128	378	60	1312	-190	473	52	16	270	90
9	420	112	493	48	1073	-256	501	44	14	194	65
10	384	330	470	80	1265	-124	562	47	15	146	49
11	693	238	447	56	1435	-119	478	54	17	245	82
12	376	211	484	50	1122	-180	563	47	15	166	55
13	595	187	496	50	1329	-181	473	45	14	224	75
14	761	83	414	51	1309	-275	389	44	14	289	96
15	504	108	455	50	1117	-226	519	34	11	203	68
16	536	234	374	69	1213	-72	636	60	19	172	57
17	702	41	385	54	1183	-319	401	32	10	272	91
18	677	28	426	50	1181	-371	350	30	9	272	91
19	787	21	389	50	1247	-403	290	42	13	311	104
20	584	32	433	50	1099	-342	408	45	14	241	80

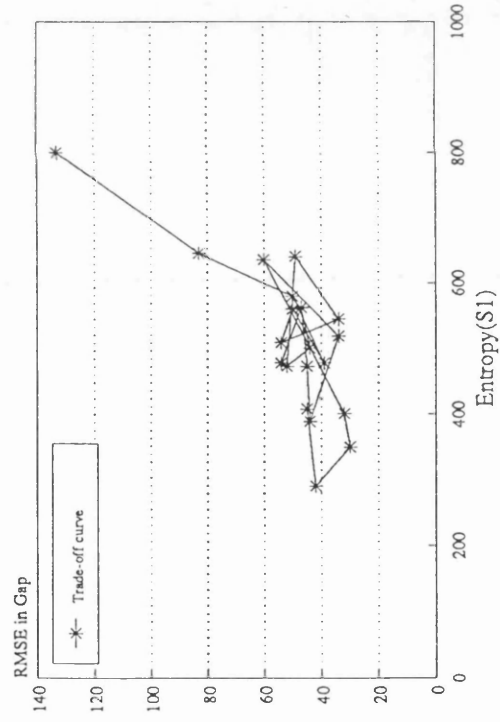
(a) Evolution of gap and goodness of fit



(b) Evolution of total demand



(d) Trade-off curve



(c) Evolution of cell values

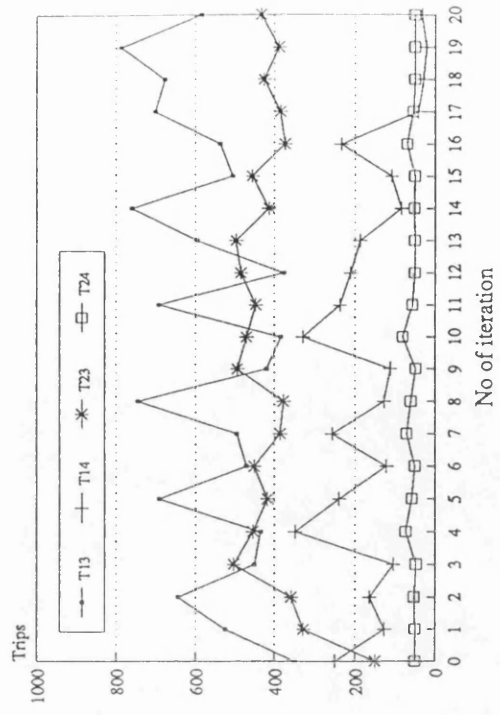


Figure 6.4 Results of the base test by ME2

However, as shown in Figures 6.4a-6.4c, an inspection of the other indicators - goodness of fit with the observed trip matrix and cell values of estimated trip matrices - shows that they are subject to considerable fluctuations throughout the 20 iterations of the test. Finally, the existence of a trade-off relationship between entropy values and gap values of estimated trip matrices was examined by plotting these two values against each other as shown in Figure 6.4d. It can be said from Figure 6.4d that it is difficult to identify any trade-off relationship after the first two iterations.

The simultaneous methods SIM-S₀ and SIM-S₁ use a sequential unconstrained optimization technique for approximating the original constrained problem. This process is accomplished by increasing the magnitude of the penalty parameter during the sequence of optimisations. The results of these tests are presented in Tables 6.3 and 6.4. The results presented in Tables 6.3 and 6.4 are depicted graphically in Figures 6.5a-6.5d and Figures 6.6a-6.6d.

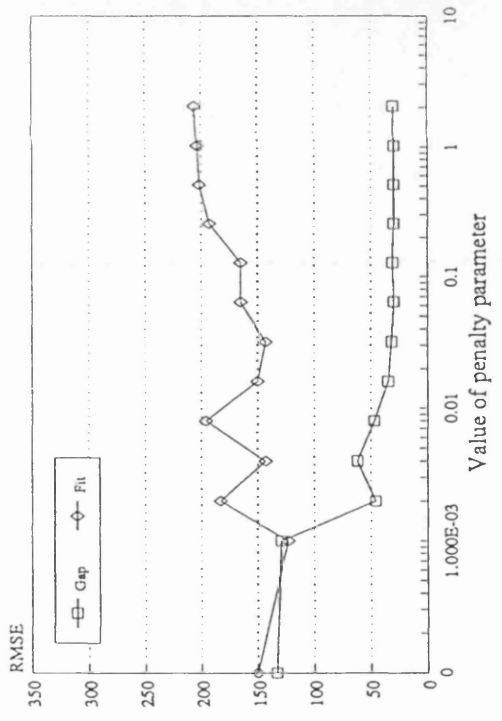
Table 6.3 Results of the base test by SIM-S₀

Penalty parameter	Trip Matrix					Entropy		Gap		Fit	
	T13	T14	T23	T24	T..	S0	S1	RMSE	%RMSE	RMSE	%RMSE
0.000	350	250	150	50	800	0	800	133	42	150	50
0.001	362	277	236	73	947	-14	773	129	40	123	41
0.002	540	457	385	90	1471	-28	547	46	14	183	61
0.004	441	369	390	78	1278	-54	625	62	20	143	48
0.008	529	469	445	72	1515	-56	491	47	15	196	65
0.016	464	224	399	83	1170	-94	631	35	11	150	50
0.032	437	223	420	93	1173	-113	611	32	10	143	48
0.064	446	196	459	73	1175	-151	572	30	10	165	55
0.128	459	193	444	72	1169	-141	585	31	10	165	55
0.256	487	182	463	30	1162	-176	552	30	10	193	64
0.512	481	205	485	5	1175	-218	506	30	9	202	67
1.024	479	210	493	1	1184	-234	486	30	9	204	68
2.048	492	193	481	1	1167	-233	493	30	9	206	69

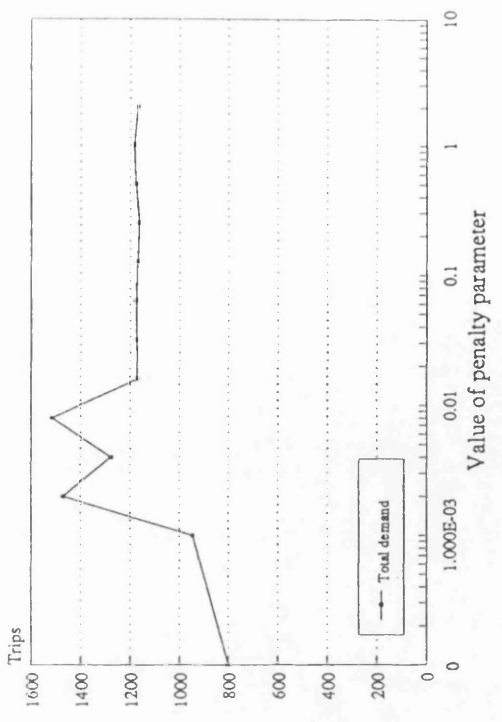
Table 6.4 Results of the base test by SIM-S₁

Penalty parameter	Trip Matrix					Entropy		Gap		Fit	
	T13	T14	T23	T24	T..	S0	S1	RMSE	%RMSE	RMSE	%RMSE
0.000	350	250	150	50	800	0	800	133	42	150	50
0.001	359	268	209	54	890	-6	789	107	33	135	45
0.002	377	245	274	56	953	-29	757	77	24	131	44
0.004	390	236	338	67	1031	-61	708	57	18	130	43
0.008	435	226	393	64	1118	-92	652	38	12	148	49
0.016	434	224	398	75	1131	-96	644	37	12	145	48
0.032	462	201	415	68	1147	-115	618	32	10	160	53
0.064	449	192	457	62	1161	-153	576	30	10	169	56
0.128	467	190	434	69	1160	-135	595	31	10	167	56
0.256	477	175	418	83	1153	-135	597	32	10	164	55
0.512	490	182	466	16	1154	-194	537	30	9	199	66
1.024	480	210	488	2	1180	-226	495	30	9	203	68
2.048	488	204	481	1	1174	-227	497	30	9	206	69

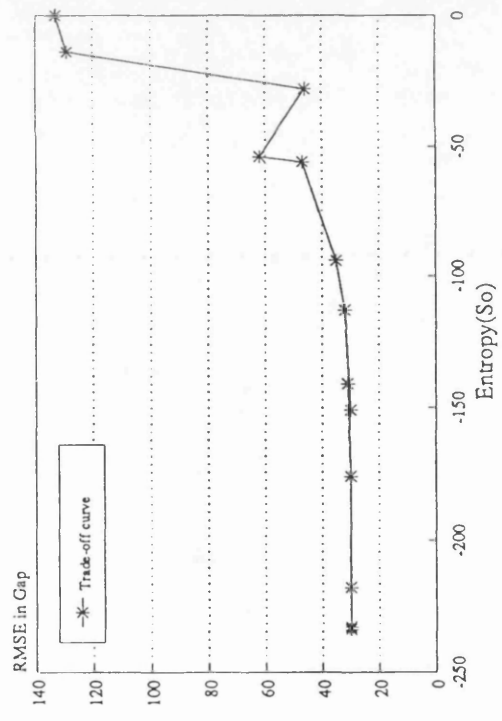
(a) Evolution of gap and goodness of fit



(b) Evolution of total demand



(d) Trade-off curve



(c) Evolution of cell values

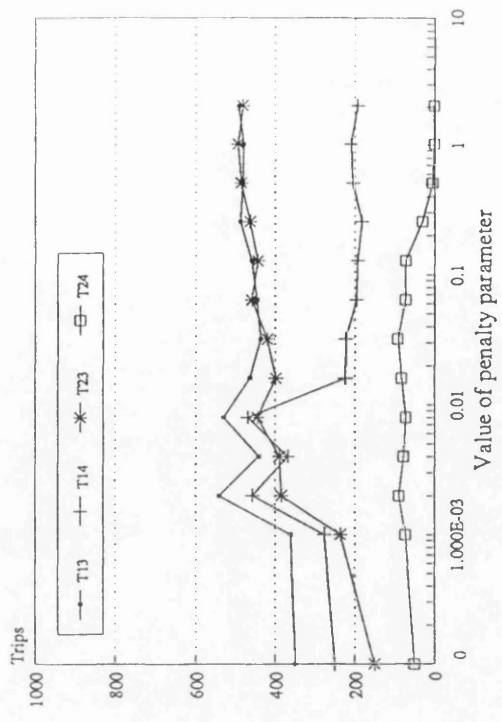
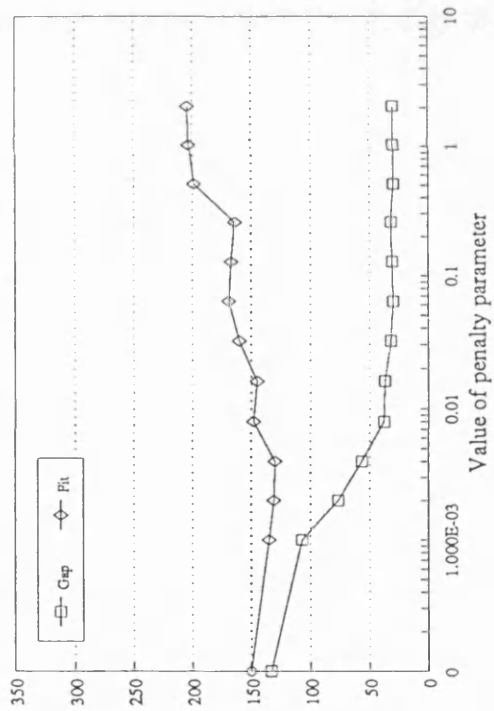
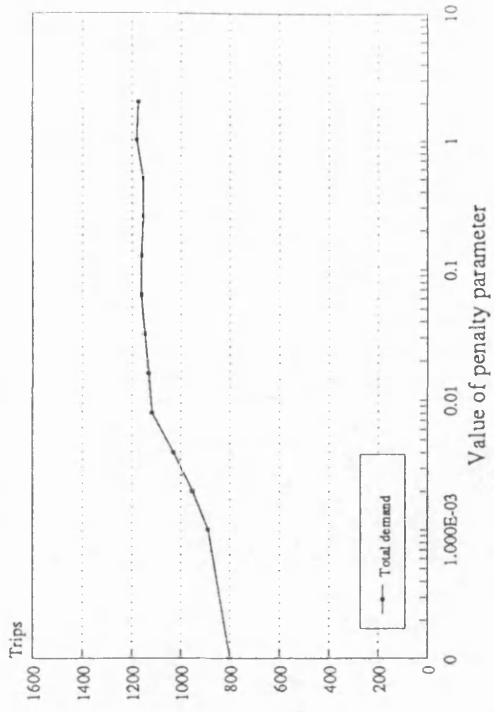


Figure 6.5 Results of the base test by SIM-So

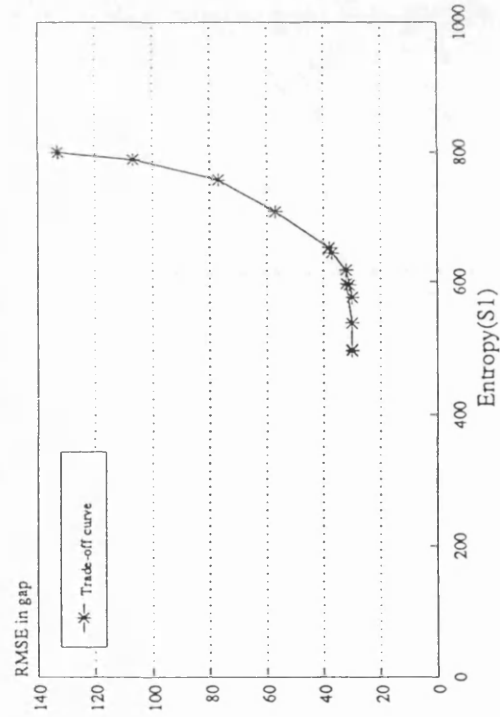
(a) Evolution of gap and goodness of fit



(b) Evolution of total demand



(d) Trade-off curve



(c) Evolution of cell values

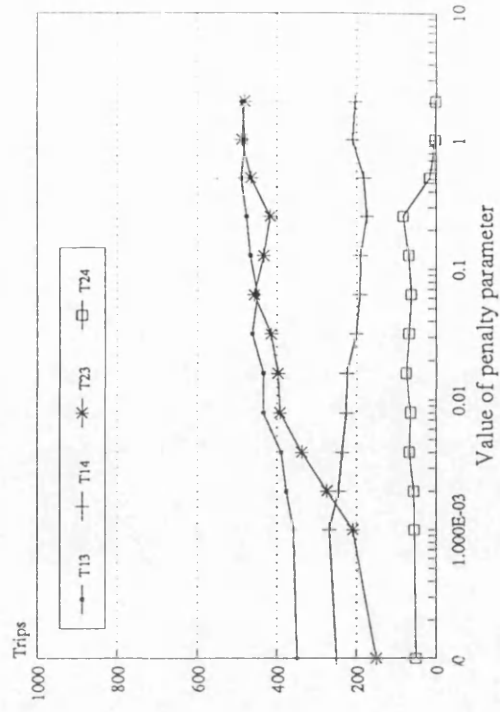


Figure 6.6 Results of the base test by SIM-S1

The following comments are made from these results.

(1) Figures 6.5a and 6.6a show that as the magnitude of the penalty parameter increases, SIM-S₀ and SIM-S₁ reduce the gap value gradually despite of the existence of a small fluctuation in the result by SIM-S₀. It is identified that SIM-S₀ fails to improve the solution in the third sub-problem. Nevertheless, both of the simultaneous methods SIM-S₀ and SIM-S₁ produce better convergence than the sequential ME2 method.

(2) Figures 6.5a and 6.6a also show change in the goodness of fit of the estimated trip matrices by SIM-S₀ and SIM-S₁. Figure 6.5a shows that SIM-S₀ improves the goodness of fit during the first iteration and thereafter the goodness of fit gets worse. On the other hand, as shown in Figure 6.6a, SIM-S₁ improves the goodness of fit during the first few sub-problems and thereafter the goodness of fit gets worse as the gap value and entropy value are further reduced.

(3) Figures 6.5b-6.5c and 6.6b-6.6c show the total demand and cell values of estimated trip matrices as the magnitude of the penalty parameter increases. It is noted that SIM-S₀ tends to overestimate the total of the estimated trip matrix during the beginning of optimisation whereas SIM-S₁ tends to underestimate. Later, both methods approach the value 1200 which is equal to the total demand of the trip matrix (300,300,300,300) which was used to generate the observed flows.

(4) Figures 6.5d and 6.6d show that a clear trade-off curve exists for each of these solution methods between gap values and entropy values. It is interesting to note that in each case after the first five or six sub-problems the entropy value is further reduced without reducing the gap. This trade-off curve could be a useful practical tool, because it allows the selection of the estimated trip matrices to be controlled depending on the relative accuracy of the prior trip matrices and observed link flows that are input.

Similar results to the simultaneous methods are obtained by the improved sequential methods ME3-S₀ and ME3-S₁. The results obtained by the improved sequential methods are presented in Tables 6.5 and 6.6. These results are depicted graphically in Figures 6.7a-6.7d and Figures 6.8a-6.8d.

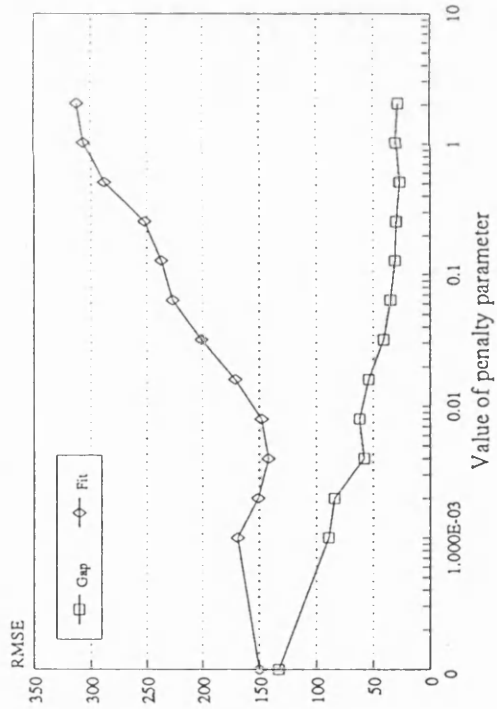
Table 6.5 Results of the base test by ME3-S₀

Penalty Parameter	Trip			Matrix		Entropy		Gap		Fit	
	T13	T14	T23	T24	T..	S0	S1	RMSE	%RMSE	RMSE	%RMSE
0.000	350	250	150	50	800	0	800	133	42	150	50
0.001	551	366	285	85	1287	-6	669	89	28	169	56
0.002	504	367	302	87	1260	-12	676	84	26	151	50
0.004	480	354	334	89	1258	-25	664	58	18	142	47
0.008	493	329	361	86	1270	-37	646	62	19	148	49
0.016	531	232	405	82	1250	-89	603	54	17	171	57
0.032	578	142	401	81	1201	-154	559	41	13	201	67
0.064	627	111	398	67	1203	-193	519	35	11	227	76
0.128	637	100	415	61	1212	-220	489	31	10	237	79
0.256	648	67	415	44	1175	-277	447	30	9	252	84
0.512	699	4	400	26	1128	-455	285	27	9	288	96
1.024	746	1	386	19	1152	-485	247	30	9	306	102
2.048	757	1	381	9	1148	-503	231	28	9	312	104

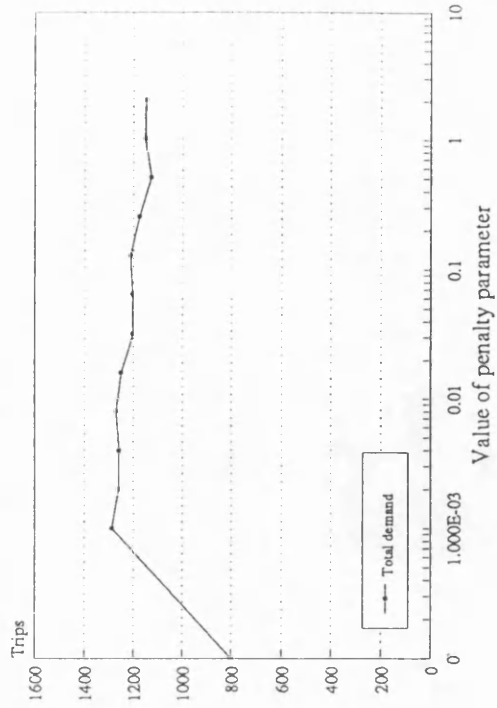
Table 6.6 Results of the base test by ME3-S₁

Penalty parameter	Trip			Matrix		Entropy		Gap		Fit	
	T13	T14	T23	T24	T..	S0	S1	RMSE	%RMSE	RMSE	%RMSE
0.000	350	250	150	50	800	0	800	133	42	150	50
0.001	397	286	215	59	957	-4	781	101	32	137	46
0.002	406	277	271	61	1015	-19	754	81	25	132	44
0.004	428	261	327	64	1080	-44	712	55	17	137	46
0.008	510	301	344	65	1221	-36	669	66	21	159	53
0.016	475	234	373	62	1144	-73	662	43	14	156	52
0.032	548	161	381	65	1155	-123	608	38	12	189	63
0.064	609	117	412	48	1186	-195	524	36	11	226	75
0.128	676	82	383	47	1189	-240	478	37	12	255	85
0.256	657	41	401	33	1131	-328	412	30	9	263	88
0.512	710	1	394	23	1128	-472	268	28	9	293	98
1.024	743	1	388	13	1144	-494	241	27	8	306	102
2.048	753	1	379	2	1135	-521	217	29	9	312	104

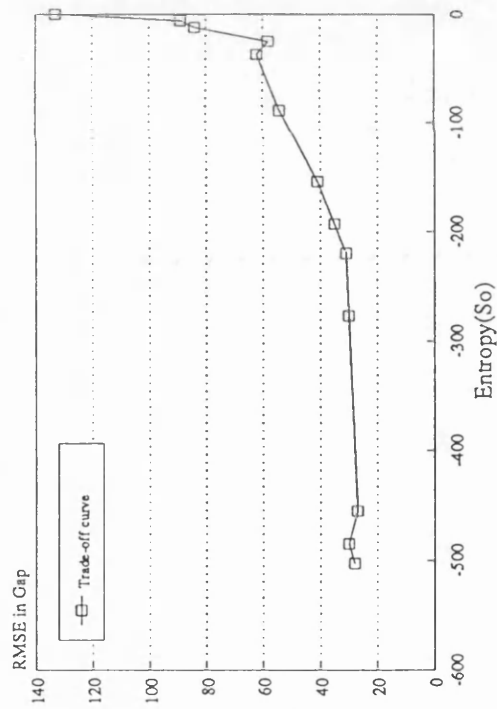
(a) Evolution of gap and goodness of fit



(b) Evolution of total demand



(d) Trade-off curve



(c) Evolution of cell values

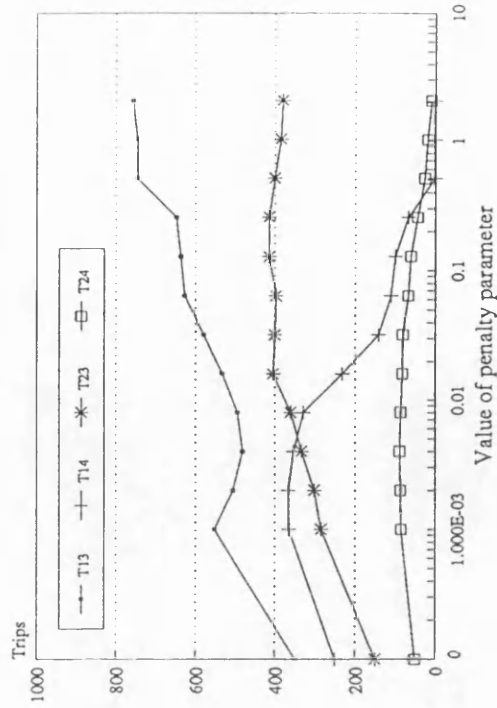


Figure 6.7 Results of the base test by ME3-S₀

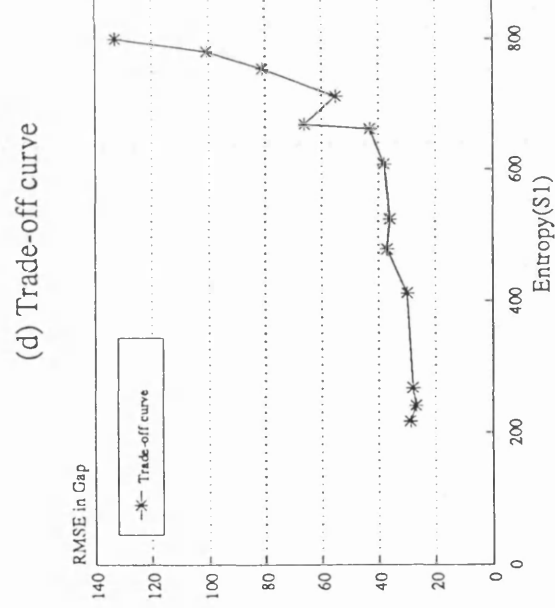
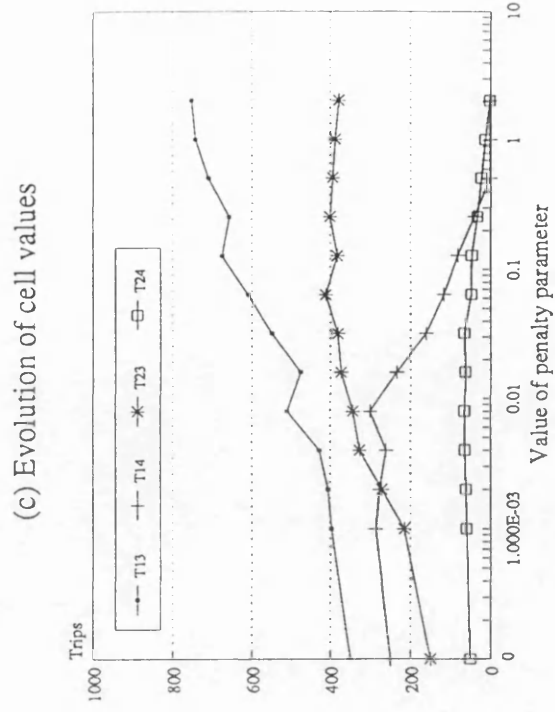
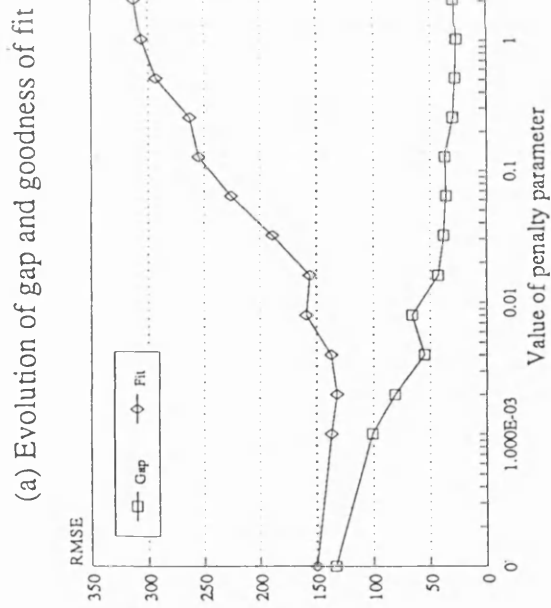
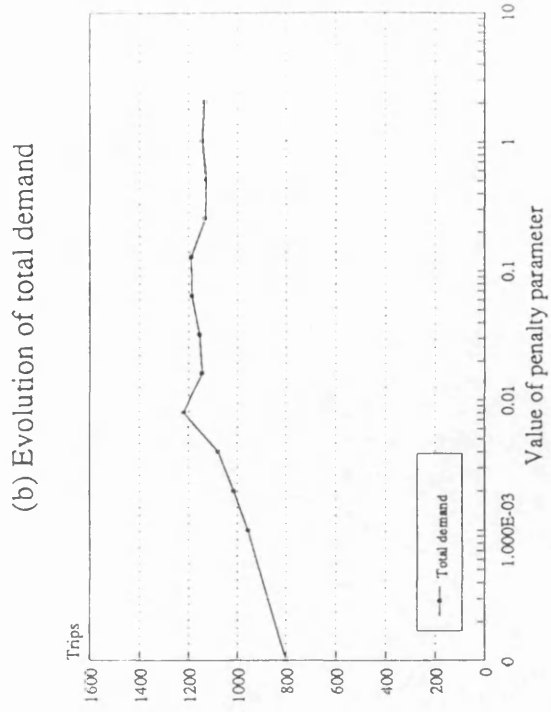


Figure 6.8 Results of the base test by ME3-S1

The following comments are made from these results.

(1) Figures 6.7a and 6.8a show that as the magnitude of the penalty parameter increases, ME3-S₀ and ME3-S₁ reduce the gap value gradually despite of the existence of a few small fluctuations. The improved sequential methods produce better convergence than the sequential ME2 method.

(2) Figure 6.8a also shows that ME3-S₁ improves the goodness of fit during the first few subproblems and thereafter the goodness of fit gets worse as the gap value and entropy value are further reduced. Figure 6.7a also shows that although ME3-S₀ fails to improve the goodness of fit during the first subproblem, a similar result to ME3-S₁ is made by ME3-S₀ after this.

(3) Figures 6.7b-6.7c and 6.8b-6.8c show the cell values of the estimated trip matrices as the magnitude of the penalty parameter increases. It is noted that, as in the case of the simultaneous methods, ME3-S₀ tends to overestimate the total of the trip matrix during the beginning of optimisation whereas ME3-S₁ tends to underestimate it. Later, both methods approach the value 1200 which is equal to the total demand of the observed trip matrix.

(4) Figures 6.7d and 6.8d show that clear trade-off curves exist between the gap values and entropy values for each of these methods.

For comparison purposes, the final results obtained by each of five estimation methods are summarized in Table 6.7 and depicted graphically in Figure 6.9. As observed in the results presented above, it is not straightforward to select a final solution for each of the competing methods because of fluctuations from iteration to iteration. Here, the final solution for the sequential ME2 method was selected at the 20th iteration and the final solutions for the improved sequential methods and the simultaneous methods were selected when convergence in gap is achieved. The performances for the estimation methods are also compared with those of the observed matrix and the prior matrix.

Table 6.7 Summary of the results of the base test by various methods

Solution Methods	Trip		Matrices			Entropy		Gap		Fit	
	T ₁₃	T ₁₄	T ₂₃	T ₂₄	T _{..}	S ₀	S ₁	RMSE	%RMSE	RMSE	%RMSE
Observed	300	300	300	300	1200	-267	446	39	12	0	0
Prior	350	250	150	50	800	0	800	133	42	150	50
ME2	584	32	433	50	1099	-342	408	45	14	241	80
ME3 - S ₁	609	117	412	48	1186	-195	524	36	11	226	75
SIM - S ₁	449	192	457	62	1161	-153	576	30	10	169	56
ME3 - S ₀	627	111	398	67	1203	-193	519	35	11	227	76
SIM - S ₀	446	196	456	73	1175	-151	572	30	10	165	55

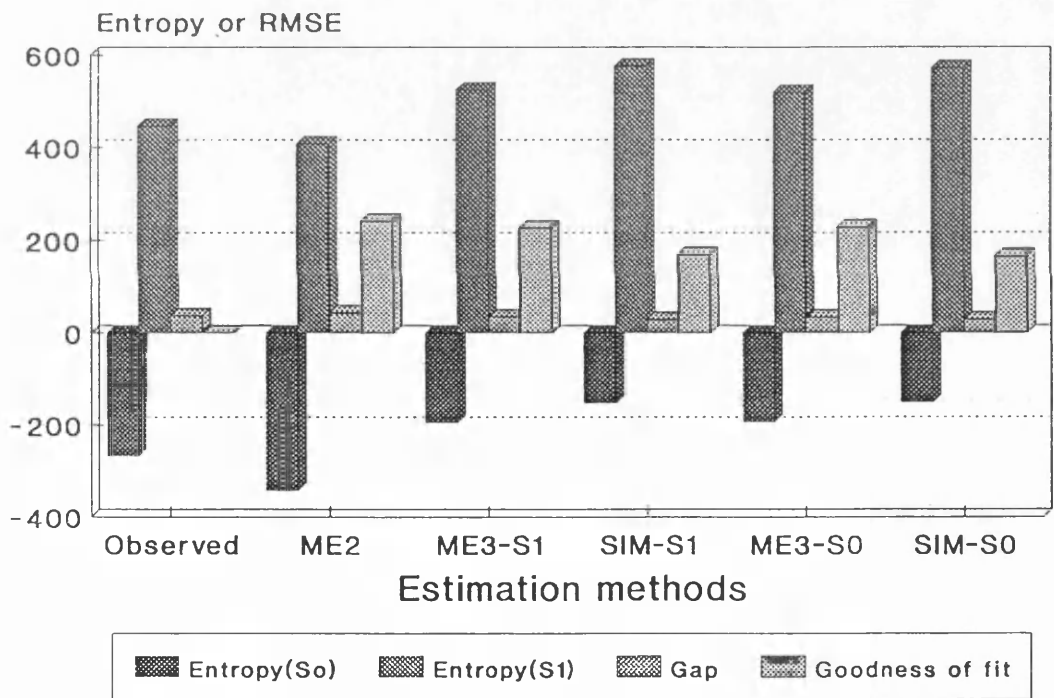


Figure 6.9 Summary of the results of the base test by various methods

It is seen from Table 6.7 and Figure 6.9 that there is no general agreement between the entropy objective values and the goodness of fit of the estimated trip matrix. The higher entropy objective values of the

simultaneous methods do not necessarily correspond to the better goodness of fit of the estimated trip matrix. The simultaneous estimation methods perform better than the sequential method ME2 and the improved sequential methods ME3-S₀ and ME3-S₁ in terms of the entropy objective values and gap values. In particular, it is interesting to note that the entropy values S₀=-342 and S₁=408 of the estimated matrices obtained by the sequential ME2 method are less than those values S₀=-267 and S₁=446 obtained from the known observed trip matrix which was used to generate the observed link flows. This indicates that the solution estimated by the sequential ME2 method is not optimal in terms of the entropy objective values. This is contrasted with the results estimated by the simultaneous methods whose entropy values are higher than those of the observed trip matrix. The solutions obtained by the simultaneous methods are better than the observed trip matrix. However, the simultaneous methods fail to improve the goodness of fit of the estimated trip matrix compared with that of the prior trip matrix, since the RMSE values of the estimated trip matrix for the simultaneous methods are higher. Perfect agreement between entropy objective values and the goodness of fit for the estimated trip matrices was not seen in this test.

Also, Table 6.7 and Figure 6.9 show that the improved sequential methods ME3-S₁ and ME3-S₀ perform better than the sequential ME2 method. This suggests that the use of the sequential estimation method combined with the penalty function approach can produce better solutions than the sequential ME2 method. This is an important observation because the improved sequential method is still practical for larger networks in terms of computing time.

It is also found from Table 6.7 that the methods using the entropy objective function $S_0(\underline{T}, t)$ tend to estimate trip matrices with totals $T_{..}$ which are greater than those estimated from the methods using the entropy objective function $S_1(\underline{T}, t)$. In this test, however, it is not possible to identify any clear difference between the solutions estimated by the methods using each of two different entropy functions.

6.3.3.3 Conclusions

It can be said that the simultaneous methods are better than the sequential ones in respect of the entropy objective values. Also, the improved sequential method seems to be a good practical alternative to the sequential ME2 method in view of the better performance and to the simultaneous method in view of the computing demand. In this test, however, there exists no good match between the entropy objective values and the goodness of fit for the estimated trip matrix. This suggests that the use of artificial data alone is not sufficient to perform a full validation test.

6.3.4 Tests with various congestion levels

6.3.4.1 Introduction

As mentioned in Section 4.4, as a road network becomes more congested, the sequential ME2 method might fail to converge or to estimate optimal solutions. As an alternative approach, the simultaneous estimation method has been proposed to solve the matrix estimation problem and equilibrium assignment problem simultaneously. It is expected that the simultaneous method will perform better than the sequential ME2 method in congested networks.

In this section, a test is designed to investigate the congestion effects on the performance of the various estimation methods. This test can be carried out effectively by estimating trip matrices from each of various sets of observed link flows whose congestion levels on the network are different. In order to generate the observed link flows with different levels of congestion, first six different trip matrices - (100,100,100,100), (200,200,200,200), (300,300,300,300), (400,400,400,400), (500,500,500,500), (700,700,700,700) - were chosen as the observed trip matrices. Each of these trip matrices was then assigned by equilibrium assignment in order to generate the associated observed link flows. Three links - (6,7), (6,10) and (9,10) - were selected for observation. The levels of congestion on links are measured by calculating values of the volume to capacity ratio. Six discrete values of mean V/C ratio obtained were such as 0.3, 0.5, 0.8, 1.0, 1.2

and 1.6. In order to remove the effects of inconsistent link flows, observed link flows were kept error-free. Finally, six non-uniform trip matrices corresponding to each of six observed trip matrices were chosen as the prior trip matrices.

The performance of the estimation methods for each of six different congestion levels was investigated by comparing the values of estimated trip matrices, entropy values and RMSE values for link flows and estimated trip matrices.

6.3.4.2 Results

The results of the tests conducted for each of six different congestion levels are summarized in Table 6.8. As shown in Table 6.8, all the estimation methods reduce the gap successfully below the level of the prior trip matrix, since the %RMSE values of the gap of the estimated trip matrices are less than 10 percent. This suggests that it is not useful to use the gap values for comparing the performance of the different estimation methods as they all perform equally well. Furthermore, there is no general agreement between the entropy objective values and the goodness of fit of the estimated trip matrices either. Therefore, it is decided that in the following analysis the values of the objective functions will be used as an index for comparing the results.

Table 6.8 Results of the tests with various congestion levels

(a) $V/C=0.3$

Solution Methods	Trip Matrices					Entropy		Gap		Fit	
	T ₁₃	T ₁₄	T ₂₃	T ₂₄	T _{..}	S ₀	S ₁	RMSE	%RMSE	RMSE	%RMSE
Observed	100	100	100	100	400	-77	166	0	0	0	0
Prior	120	80	50	20	270	0	270	39	39	49	49
ME2	101	133	132	30	397	-39	205	0	0	42	42
ME3 - S ₁	102	133	133	30	397	-39	205	0	0	42	42
SIM - S ₁	123	92	106	70	390	-38	208	7	7	20	20
ME3 - S ₀	100	125	130	36	391	-38	208	1	1	38	38
SIM - S ₀	122	92	103	82	399	-49	194	5	5	15	15

Table 6.8 (cont.)

(b) $V/C=0.5$

Solution Methods	Trip		Matrices			Entropy		Gap		Fit	
	T ₁₃	T ₁₄	T ₂₃	T ₂₄	T _{..}	S ₀	S ₁	RMSE	%RMSE	RMSE	%RMSE
Observed	200	200	200	200	800	-193	277	0	0	0	0
Prior	230	170	100	30	530	0	530	80	44	101	50
ME2	58	173	327	162	720	-363	136	12	7	98	49
ME3 - S ₁	62	197	327	144	729	-328	169	9	5	98	49
SIM - S ₁	185	210	223	191	810	-194	272	4	2	15	8
ME3 - S ₀	52	200	330	158	740	-365	128	6	3	101	50
SIM - S ₀	198	231	207	160	796	-136	337	9	5	26	13

(c) $V/C=0.8$

Solution Methods	Trip		Matrices			Entropy		Gap		Fit	
	T ₁₃	T ₁₄	T ₂₃	T ₂₄	T _{..}	S ₀	S ₁	RMSE	%RMSE	RMSE	%RMSE
Observed	300	300	300	300	1200	-267	446	0	0	0	0
Prior	350	250	150	50	800	0	800	119	45	150	50
ME2	103	199	376	403	1081	-691	65	9	3	128	43
ME3 - S ₁	52	349	307	476	1184	-845	-125	5	2	154	51
SIM - S ₁	460	225	445	84	1213	-123	585	6	2	157	52
ME3 - S ₀	450	263	440	102	1255	-108	582	5	2	144	48
SIM - S ₀	503	347	497	89	1436	-103	493	8	3	178	59

(d) $V/C=1.0$

Solution Methods	Trip		Matrices			Entropy		Gap		Fit	
	T ₁₃	T ₁₄	T ₂₃	T ₂₄	T _{..}	S ₀	S ₁	RMSE	%RMSE	RMSE	%RMSE
Observed	400	400	400	400	1600	-356	596	0	0	0	0
Prior	470	330	200	67	1067	0	1067	132	53	200	50
ME2	378	528	598	116	1620	-208	736	6	2	185	46
ME3 - S ₁	660	509	495	77	1742	-51	837	4	1	220	55
SIM - S ₁	727	283	518	48	1576	-136	825	2	1	254	64
ME3 - S ₀	626	578	435	104	1742	-32	856	3	1	207	52
SIM - S ₀	671	526	466	107	1770	-33	842	5	2	212	53

(e) $V/C=1.2$

Solution Methods	Trip		Matrices			Entropy		Gap		Fit	
	T ₁₃	T ₁₄	T ₂₃	T ₂₄	T _{..}	S ₀	S ₁	RMSE	%RMSE	RMSE	%RMSE
Observed	500	500	500	500	2000	-460	724	0	0	0	0
Prior	580	420	250	80	1330	0	1330	150	38	251	50
ME2	346	443	571	463	1823	-555	693	31	8	92	18
ME3 - S ₁	610	588	688	93	1980	-153	1040	20	5	235	47
SIM - S ₁	864	492	662	73	2091	-114	1031	13	3	292	58
ME3 - S ₀	742	541	637	145	2064	-93	1064	12	3	226	45
SIM - S ₀	774	548	660	100	2081	-99	1050	10	3	256	51

(f) $V/C=1.6$

Solution Methods	Trip		Matrices			Entropy		Gap		Fit	
	T ₁₃	T ₁₄	T ₂₃	T ₂₄	T _{..}	S ₀	S ₁	RMSE	%RMSE	RMSE	%RMSE
Observed	700	700	700	700	2800	-610	1059	0	0	0	0
Prior	820	580	350	120	1087	0	1870	213	37	349	50
ME2	511	923	1001	191	2627	-436	1298	27	5	330	47
ME3 - S ₁	883	867	845	138	2733	-141	1555	23	4	315	45
SIM - S ₁	1115	735	752	161	2763	-61	1624	18	4	341	49
ME3 - S ₀	1015	806	822	196	2838	-94	1560	13	3	308	44
SIM - S ₀	1116	787	775	178	2855	-61	1586	13	3	339	48

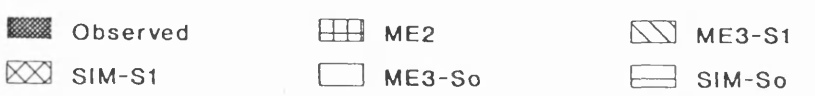
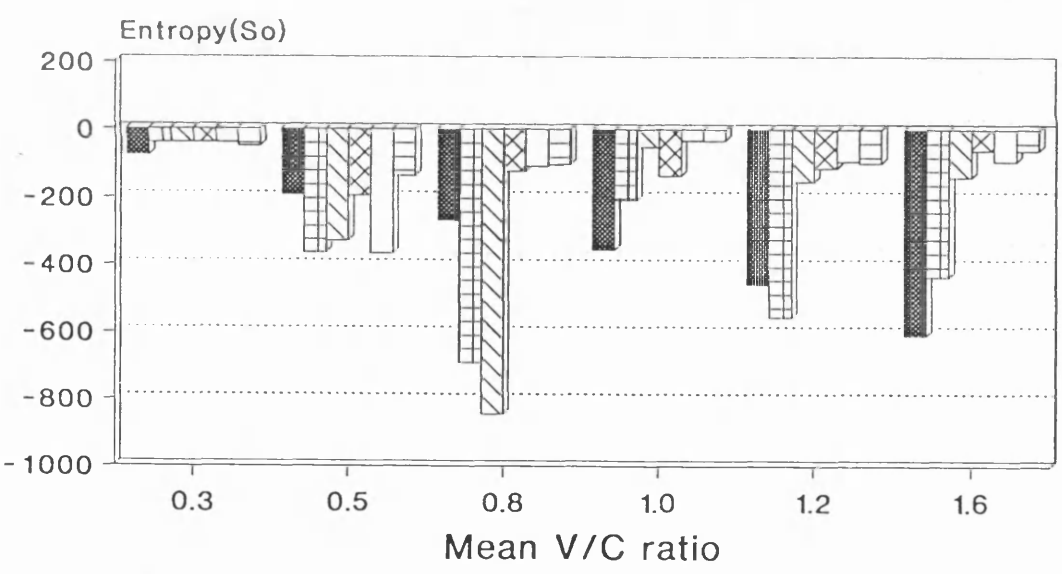
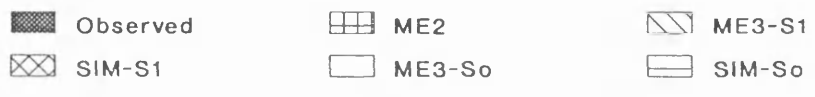
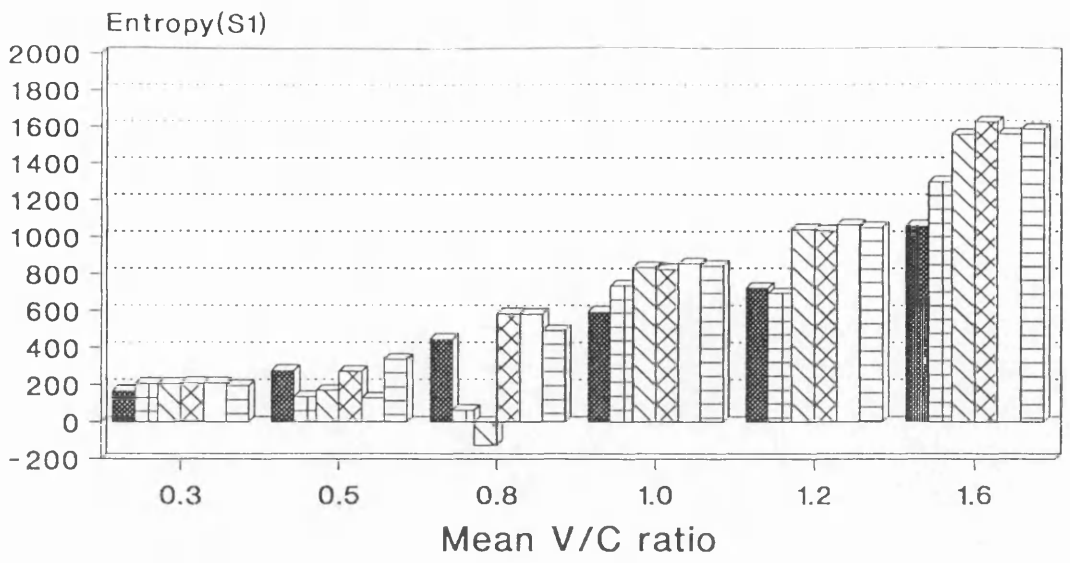


Figure 6.10 Results of the tests with various congestion levels

It can be seen from Table 6.8 and Figure 6.10 that the sequential ME2 method produces good results at low congestion levels ($V/C=0.3$) and at the high congestion levels ($V/C=1.0$ or 1.6) compared with the results by the improved sequential methods. The sequential ME2 method produces its worst results at the congestion level of $V/C=0.8$. The entropy values of the trip matrices estimated by the sequential ME2 method are higher than the observed one at the congestion levels $V/C=0.3$, 1.0 or 1.6 . However, at the congestion levels $V/C=0.5$, 0.8 and 1.2 , the sequential ME2 method estimates trip matrices having lower entropy values than those of the observed trip matrices.

The improved sequential methods ME3-S₁ and ME3-S₀ produce similar results to the sequential ME2 method at lower levels of V/C but perform much better at higher levels. It is interesting to observe that the improved sequential methods produce the same solution at low congestion level $V/C=0.3$ as does the ME2. However, the improved sequential ME3-S₁ method produces the worst result of all the five estimation methods at the congestion level $V/C=0.8$, although it produces better results than the sequential ME2 method at other congestion levels, especially at higher levels.

The simultaneous methods SIM-S₁ and SIM-S₀ produce the best results of all the estimation methods. Also, the simultaneous methods estimate trip matrices having higher entropy values than the observed ones in all the cases, although perfect feasibility is not obtained for the solutions.

In this test, as shown in Table 6.8, it is difficult to identify any particular differences between the methods on the basis of the results from the use of different entropy objective functions. However, as observed from the results of the base test in Section 6.3.3, the solutions estimated from the use of the entropy objective function $S_0(\underline{T}, \underline{t})$ tend to have greater total trip demand than those from the use of the entropy objective function $S_1(\underline{T}, \underline{t})$.

6.3.4.3 Conclusions

The results of the tests with various congestion levels show that the sequential ME2 method tends to have a difficulty to produce good results in moderately congested networks compared with the simultaneous methods. On the other hand, the simultaneous methods produce the best solutions of all the estimation methods consistently at any congestion levels.

6.3.5 Tests with inconsistent link flows

6.3.5.1 Introduction

The use of mutually inconsistent traffic counts might lead to there being no feasible solution in the matrix estimation problem. Two sources of inconsistency in traffic counts were identified in Chapter 3. The first one is that errors in the counts may lead to situations in which the total flow into a node is not equal to the total flow out of the same node, thus violating the flow conservation conditions. The second source is due to a mismatch between the assumed traffic assignment model and the observed flows. This type of inconsistency occurs whenever path flow continuity is not met. The fact that path flow continuity conditions are not met seems to reflect errors in assignment whereas the link flow discontinuities are a reflection solely of errors in the traffic counts.

The proposed simultaneous and improved sequential estimation methods are expected to be more flexible in dealing with inconsistency in link flows than the sequential ME2 method. This is due to the use of the quadratic gap function and the penalty function approach in which full feasibility is not necessarily required. Therefore, it is interesting to investigate any differences in the performance of the estimation methods when traffic counts are not error-free.

In this section, a test is designed to investigate the effects of inconsistent link flows on the performance of various estimation methods. This test was accomplished by estimating trip matrices from

inconsistent link flows. By assigning a known trip matrix (300,300,300,300), the flows on some particular links are obtained. Four links - (6,7), (6,10), (9,10) and (10,11) - were selected for observation. Artificial errors having a rectangular distribution were then added to these assigned link flows. Five error dispersions in link flows were selected. These six levels were within a range of $\pm 0\%$, $\pm 10\%$, $\pm 20\%$, $\pm 30\%$ and $\pm 50\%$ of the mean flows. Finally, a trip matrix (350,250,150,50) which is non-uniform to the observed trip matrix was chosen as a prior trip matrix.

The performance of the estimation methods for each of five different error levels is investigated by comparing the values of estimated trip matrices, entropy values, gap values and the goodness of fit of the estimated trip matrices.

6.3.5.2 Results

It is interesting to see how the estimation methods behave as the level of error in link flows increases. Figures 6.11a-e show the reduction of gap by various estimation methods.

(a) ME2

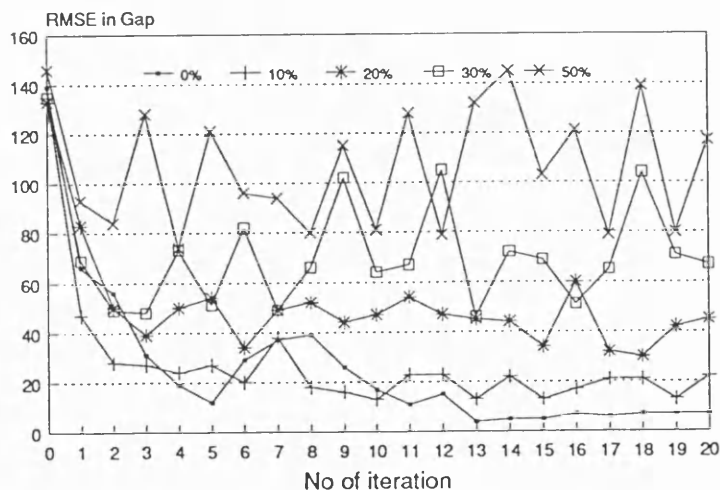
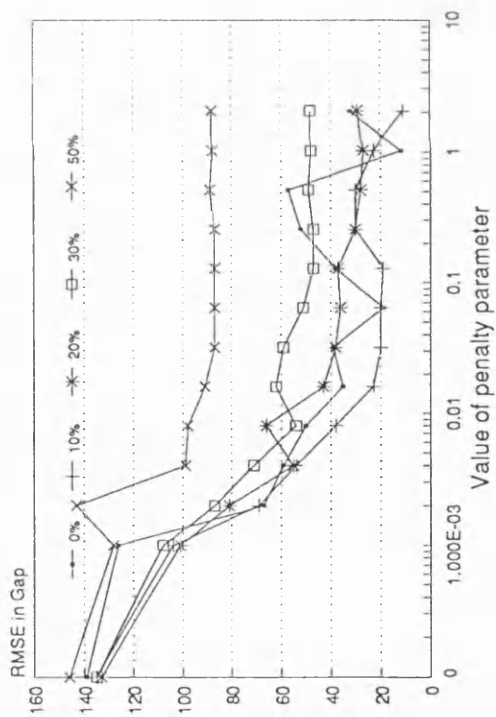
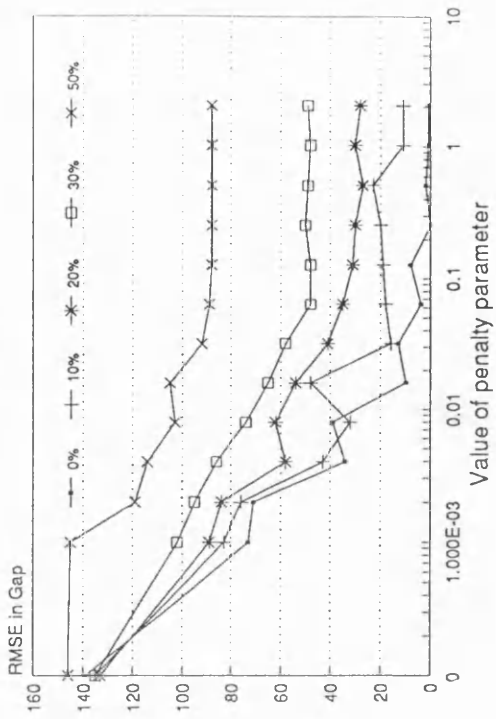


Figure 6.11 Results of the tests with inconsistent link flows

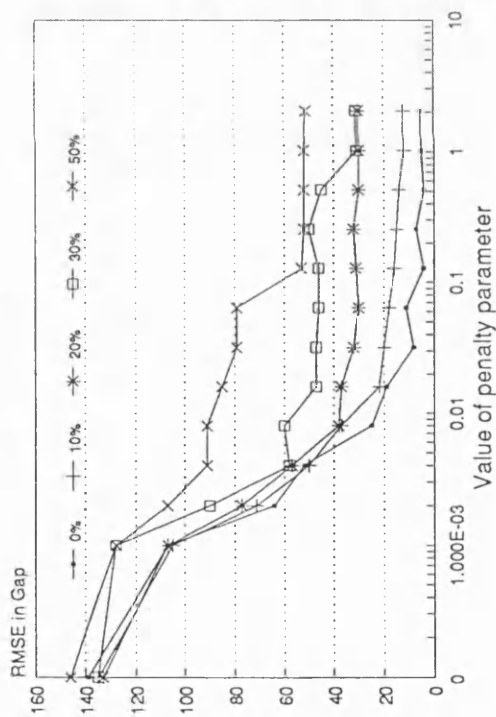
(b) ME3-S1



(d) ME3-So



(c) SIM-S1



(e) SIM-So

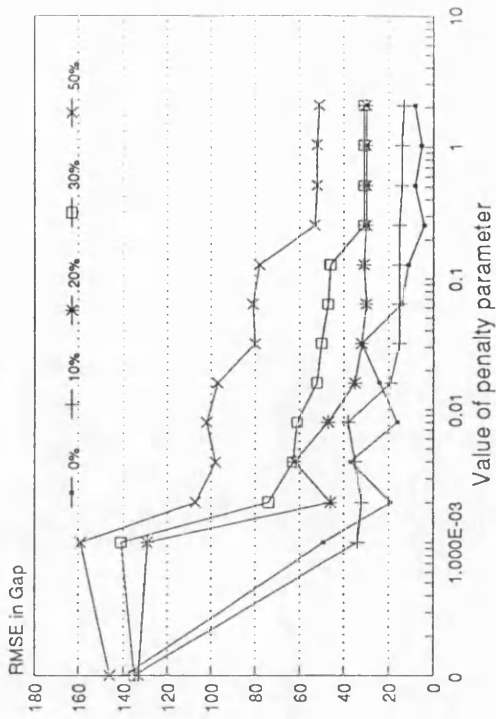


Figure 6.11 (cont.) Results of the tests with inconsistent link flows

Figure 6.11a shows the reduction of gap by the sequential ME2 method as the number of iterations increases. Figure 6.11a shows that the sequential ME2 method oscillates more as the level of error in link flows increases. On the other hand, as shown in Figures 6.11c and 6.11e, the simultaneous methods SIM-S₁ and SIM-S₀ tend to converge well compared with the sequential ME2 method, as the magnitude of the penalty parameter increases. Also, as shown in Figures 6.11b and 6.11d, the improved sequential methods ME3-S₁ and ME3-S₀ perform better than the sequential ME2 method, although there exist small fluctuations.

The final results of the tests are summarized in Table 6.9. It can be seen from Table 6.9 that the simultaneous methods perform best in terms of gap and entropy values. The improved sequential methods perform better than the sequential ME2 method, but it tends to be more unstable compared with the simultaneous methods. It can be also observed that the sequential ME2 method fails to reduce the gap value, as the level of error in link flows increases.

It can be seen from Figures 6.11c and 6.11e that the simultaneous method SIM-S₁ performs better at the early period of the sequence than the simultaneous method SIM-S₀. At the end, both of the methods produce similar results. In the case of the improved sequential methods, as shown in Figures 6.11b and 6.11d, ME3-S₀ performs better than ME3-S₁. However, it is not possible to identify any significant difference between the results due to the use of different entropy functions $S_0(T,t)$ and $S_1(T,t)$.

Table 6.9 Results of the tests with inconsistent link flows

(a) 0%

Solution Methods	Trip		Matrices			Entropy		Gap		Fit	
	T ₁₃	T ₁₄	T ₂₃	T ₂₄	T _{..}	S ₀	S ₁	RMSE	%RMSE	RMSE	%RMSE
Observed	300	300	300	300	1200	-267	446	0	0	0	0
Prior	350	250	150	50	800	0	800	139	44	150	50
ME2	90	225	360	433	1109	-745	2	7	2	134	45
ME3 - S ₁	262	245	595	83	1185	-316	404	19	6	186	62
SIM - S ₁	469	235	442	91	1237	-116	582	4	1	156	52
ME3 - S ₀	464	266	437	101	1268	-102	582	4	1	147	49
SIM - S ₀	481	263	468	72	1284	-117	559	4	1	169	56

Table 6.9 (cont.)

(b) 10%

Solution Methods	Trip		Matrices		T..	Entropy		Gap		Fit	
	T ₁₃	T ₁₄	T ₂₃	T ₂₄		S ₀	S ₁	RMSE	%RMSE	RMSE	%RMSE
Observed	300	300	300	300	1200	-267	446	19	6	0	0
Prior	350	250	150	50	800	0	800	135	41	150	50
ME2	663	25	379	48	1116	-345	400	22	7	263	88
ME3 - S ₁	509	201	426	66	1202	-121	592	20	6	176	59
SIM - S ₁	486	207	400	81	1175	-102	622	18	6	159	53
ME3 - S ₀	505	183	435	84	1208	-139	572	18	6	174	58
SIM - S ₀	460	233	440	86	1219	-116	590	15	5	154	51

(c) 20%

Solution Methods	Trip		Matrices		T..	Entropy		Gap		Fit	
	T ₁₃	T ₁₄	T ₂₃	T ₂₄		S ₀	S ₁	RMSE	%RMSE	RMSE	%RMSE
Observed	300	300	300	300	1200	-267	446	39	12	0	0
Prior	350	250	150	50	800	0	800	133	42	150	50
ME2	584	32	433	50	1099	-342	408	45	14	241	80
ME3 - S ₁	609	117	412	48	1186	-195	524	36	11	226	75
SIM - S ₁	449	192	457	62	1161	-153	576	30	10	169	56
ME3 - S ₀	627	111	398	67	1203	-193	519	35	11	227	76
SIM - S ₀	446	196	456	73	1175	-151	572	30	10	165	55

(d) 30%

Solution Methods	Trip		Matrices		T..	Entropy		Gap		Fit	
	T ₁₃	T ₁₄	T ₂₃	T ₂₄		S ₀	S ₁	RMSE	%RMSE	RMSE	%RMSE
Observed	300	300	300	300	1200	-267	446	58	18	0	0
Prior	350	250	150	50	800	0	800	135	42	150	50
ME2	532	152	583	50	1317	-282	379	67	21	234	78
ME3 - S ₁	634	79	386	44	1143	-237	499	51	16	242	81
SIM - S ₁	457	168	472	41	1137	-188	549	46	14	186	62
ME3 - S ₀	566	145	401	71	1183	-150	570	58	18	199	66
SIM - S ₀	447	202	375	87	1111	-93	653	52	16	143	48

(e) 50%

Solution Methods	Trip		Matrices		T..	Entropy		Gap		Fit	
	T ₁₃	T ₁₄	T ₂₃	T ₂₄		S ₀	S ₁	RMSE	%RMSE	RMSE	%RMSE
Observed	300	300	300	300	1200	-267	446	97	31	0	0
Prior	350	250	150	50	800	0	800	146	47	150	50
ME2	671	349	457	50	1527	-75	464	117	37	238	79
ME3 - S ₁	620	81	366	38	1104	-223	525	91	29	237	79
SIM - S ₁	436	122	492	28	1079	-255	502	79	25	201	67
ME3 - S ₀	609	117	370	47	1144	-172	563	92	29	223	74
SIM - S ₀	392	169	423	80	1064	-151	609	80	25	149	50

6.3.5.3 Conclusions

The results show that the simultaneous methods and the improved sequential methods are less disrupted than the sequential ME2 method by the existence of errors in link flows. Furthermore, the sequential ME2 method cannot accommodate errors in link flows and fails to converge to a stable solution when the error range exceeds 30% of the mean link flows.

6.3.6 Tests with determinate matrix total

6.3.6.1 Introduction

The problem of estimating a trip matrix from traffic counts is underspecified in most cases. Furthermore, when the estimation problem is solved under equilibrium traffic conditions, the estimation problem is further underspecified because the values of the route choice proportions in the equilibrium assignment are not defined uniquely. The entropy objective functions are used to determine a trip matrix from many feasible trip matrices.

In Chapter 5, two different entropy objective functions $S_1(\underline{T}, \underline{t})$ and $S_0(\underline{T}, \underline{t})$ were introduced. It was noted that the entropy function $S_1(\underline{T}, \underline{t})$ is equivalent to $S_0(\underline{T}, \underline{t})$ if the total demand $T_{..}$ of the estimated trip matrix is fixed. However, in practice the total demand is unknown and it should be treated as a variable rather than a constant. It was revealed in Section 4.3.2 that the use of the entropy function $S_1(\underline{T}, \underline{t})$ leads to estimation of the incorrect solutions in some cases. As an alternative, the full entropy function $S_0(\underline{T}, \underline{t})$ in which the total demand $T_{..}$ is not a constant has been proposed. At this stage, it is interesting to investigate whether there are any differences on the performance between these two entropy objective functions.

In this section, a test is designed to investigate the effects on the performance due to the use of two different entropy functions. This test can be accomplished by comparing the results from tests of using two different sets of observed link flows. The first test is to estimate a trip matrix from observed link flows from which the total demand of the estimated trip matrix is not determined. The second one is to estimate a trip matrix from observed link flows from which the total demand of the estimated trip matrix is known prior to the matrix estimation. For example, the data for the second test can be obtained by observing the flows of links located on a screen line of the network. In this test, for the first test, three links - (6,7), (6,10), and (9,10) - were selected for observation and for the second test five links - (12,8), (6,7), (6,10), (9,10) and (9,13) - were observed.

6.3.6.2 Results

The results of the tests are summarized in Tables 6.10a-b. It can be seen from this table that the totals of all of the trip matrices estimated with a determinate matrix total tend to be very close to the given total $T_{..}=1200$ compared with the results estimated with indeterminate matrix total. Table 6.10a shows that use of either entropy functions fails to estimate a trip matrix whose total is equal to the total of the observed trip matrix. In particular, a comparison of the results of two tests indicates that the simplified entropy function $S_1(T, t)$ tends to estimate trip matrices with totals that are lower than the total of the observed matrix. On the other hand, the full entropy function $S_0(T, t)$ tends to estimate trip matrices with totals that are greater than the totals of the observed matrix.

Table 6.10 Results of the tests with determinate matrix total

(a) with undetermined matrix total

Solution Methods	Trip Matrices					Entropy		Gap		Fit	
	T ₁₃	T ₁₄	T ₂₃	T ₂₄	T _{..}	S ₀	S ₁	RMSE	%RMSE	RMSE	%RMSE
Observed	300	300	300	300	1200	-267	446	0	0	0	0
Prior	350	250	150	50	800	0	800	119	47	150	50
ME2	103	199	376	403	1081	-691	65	9	3	128	43
ME3 - S ₁	52	349	307	476	1184	-845	-125	5	2	154	51
SIM - S ₁	460	225	445	84	1213	-123	585	6	2	157	52
ME3 - S ₀	450	263	440	102	1255	-108	582	5	2	144	48
SIM - S ₀	503	347	497	89	1436	-103	493	8	3	178	59

(b) with determinate matrix total

Solution Methods	Trip Matrices					Entropy		Gap		Fit	
	T ₁₃	T ₁₄	T ₂₃	T ₂₄	T _{..}	S ₀	S ₁	RMSE	%RMSE	RMSE	%RMSE
Observed	300	300	300	300	1200	-267	446	0	0	0	0
Prior	350	250	150	50	800	0	800	109	46	150	50
ME2	389	44	349	427	1209	-676	34	11	4	152	51
ME3 - S ₁	308	86	420	385	1200	-604	110	8	3	129	43
SIM - S ₁	366	203	315	311	1195	-296	420	19	8	59	20
ME3 - S ₀	281	99	456	365	1201	-591	122	7	3	132	44
SIM - S ₀	354	235	370	247	1207	-223	488	27	11	61	20

It is interesting to note from Table 6.10b that there is a good agreement between entropy objective values and the goodness of fit of the estimated trip matrices. That is, the estimated trip matrices which have higher entropy values also have better goodness of fit between the estimated matrix and the observed one. This good match is obtained because the estimation problem in the second test is more constrained by allowing more links to be observed than the first test.

It can be said from Tables 6.10a-b that the simultaneous estimation methods produce better performance than either the sequential ME2 method or the improved sequential ME3 methods, although the gap values for the trip matrices estimated from the simultaneous method are slightly worse than those from ME2 or ME3.

6.3.6.3 Conclusions

Both of the entropy objective functions $S_1(\underline{T}, \underline{t})$ and $S_0(\underline{T}, \underline{t})$ fail to estimate a trip matrix whose total is equal to the total of the observed trip matrix when the estimation problem is underspecified and the information of the total demand is not determined from observed link flows. The results show that use of the entropy objective function $S_1(\underline{T}, \underline{t})$ tends to underestimate the totals of the estimated trip matrices whereas use of the entropy objective function $S_0(\underline{T}, \underline{t})$ tends to overestimate the totals. Use of observed link flows which determine the matrix total reduces the difference between the matrices estimated using each of the objective functions.

6.3.7 Tests with different prior trip matrices

6.3.7.1 Introduction

The estimation problem can be interpreted as a process of finding a new trip matrix which is as similar as possible to the prior trip matrix and which reproduces observed link flows. Thus, the prior trip matrix plays an important role for estimating a trip matrix from observed link flows. This is particularly so when, as is usual, the estimation problem is underspecified. Furthermore, when trip matrices are estimated subject

to equilibrium traffic constraints, in the case of the sequential estimation method the prior trip matrix is used to generate initial route choice proportions for the matrix estimation process. In the case of the simultaneous method, the prior trip matrix provides an initial search interval from which the solutions are found.

In this section, a test is designed to investigate any effects on the performance of the estimation methods due to the quality of the prior trip matrix. In this test, the quality of the prior trip matrix is judged by the difference between the prior trip matrix and the one used to generate observed link flows, although this is not possible in real problems. Three different prior trip matrices t were used in the test. The first prior trip matrix - (200,200,200,200) - was uniformly scaled from the observed trip matrix (300,300,300,300). The second one - (260,230,340,280) - was selected to be similar to the observed one. The third one - (350,250,150,50) - was chosen to have a different trip distribution pattern.

6.3.7.2 Results

The results of the tests for each of three prior trip matrices are shown in Tables 6.11a-c. In the case of the first prior trip matrix, it is possible to know the optimal solution to the methods using the entropy function $S_o(T,t)$ before matrix estimation, which should be equal to the observed trip matrix. This can be used to check the optimality of the solutions estimated by these estimation methods.

Table 6.11 Results of the tests with different prior trip matrices

(a) with prior trip matrix uniformly scaled from the observed matrix

Solution Methods	Trip Matrices					Entropy		Gap		Fit	
	T13	T14	T23	T24	T..	S _o	S ₁	RMSE	%RMSE	RMSE	%RMSE
Observed	300	300	300	300	1200	0	713	0	0	0	0
Prior	200	200	200	200	800	0	800	84	35	100	33
ME2	150	182	358	444	1135	-106	632	6	2	123	41
ME3 - S ₁	173	142	455	358	1128	-121	619	12	5	131	44
SIM - S ₁	419	248	373	226	1266	-43	643	6	2	83	28
ME3 - S _o	307	307	307	307	1229	0	701	3	1	7	2
SIM - S _o	335	330	322	325	1311	0	663	6	2	28	9

Table 6.11 (cont.)

(b) with prior trip matrix close to the observed trip matrix

Solution Methods	Trip Matrices					Entropy		Gap		Fit	
	T ₁₃	T ₁₄	T ₂₃	T ₂₄	T _{..}	S ₀	S ₁	RMSE	%RMSE	RMSE	%RMSE
Observed	300	300	300	300	1200	-12	1094	0	0	0	0
Prior	260	230	340	280	1110	0	1110	67	22	46	15
ME2	159	199	328	459	1145	-73	1037	1	0	119	40
ME3 - S ₁	206	135	465	354	1161	-57	1051	9	3	129	43
SIM - S ₁	431	221	388	212	1251	-51	1050	3	1	99	33
ME3 - S ₀	272	205	414	339	1229	-9	1095	9	3	78	26
SIM - S ₀	318	270	397	338	1322	-1	1090	12	5	60	20

(c) with prior trip matrix not uniformly scaled from the observed matrix

Solution Methods	Trip Matrices					Entropy		Gap		Fit	
	T ₁₃	T ₁₄	T ₂₃	T ₂₄	T _{..}	S ₀	S ₁	RMSE	%RMSE	RMSE	%RMSE
Observed	300	300	300	300	1200	-267	446	0	0	0	0
Prior	350	250	150	50	800	0	800	119	45	150	50
ME2	103	199	376	403	1081	-691	65	9	3	128	43
ME3 - S ₁	52	349	307	476	1184	-845	-125	5	2	154	51
SIM - S ₁	460	225	445	84	1213	-123	585	6	2	157	52
ME3 - S ₀	450	263	440	102	1255	-108	582	5	2	144	48
SIM - S ₀	503	347	497	89	1436	-103	493	8	3	178	59

It can be seen from Table 6.11a that both of the simultaneous methods SIM-S₀ and SIM-S₁ estimate the trip matrices which are close to the observed trip matrix, although the estimated trip matrices are not exactly equal to the observed one. In this test, the estimated trip matrix having higher entropy values also produces better goodness of fit between the estimated matrix and the observed one. Also, Table 6.11a shows that the estimation methods using the entropy function $S_0(\underline{T}, t)$ perform better than the methods using $S_1(\underline{T}, t)$.

Tables 6.11a-c show that the trip matrices estimated by the sequential methods ME2 or ME3 are not much affected from the different prior trip matrices compared with the results estimated by the simultaneous estimation methods, although the use of better prior trip matrices produces estimated trip matrices with higher entropy objective values. The sequential method ME2 and the improved sequential methods ME3-S₀ and ME3-S₁ tend to estimate trip matrices with a similar distribution pattern to each other. By contrast, in the case of the simultaneous methods, use of prior trip matrices closer to the observed matrix leads to the estimation of trip matrices closer to the observed matrix.

6.3.7.3 Conclusions

It can be said from these results that the sequential method ME2 and the improved sequential methods ME3-S₀ and ME3-S₁ fail to reflect the information of the prior trip matrices in estimating trip matrices. This suggests that the sequential method and the improved sequential methods tend to find a feasible solution rather than an optimal one. By contrast, the simultaneous estimation methods SIM-S₀ and SIM-S₁ tend to be more faithful to the information of the prior trip matrices.

6.3.8 Tests with different sets of links observed

6.3.8.1 Introduction

In Section 4.4.1, analysis of a simple example showed that selection of links observed for traffic counts can have a profound influence on the convergence of the sequential method ME2. It was suggested that the sequential method ME2 converges well when the links which have the flows loaded in the first (uncongested) all or nothing assignment are observed for matrix estimation. On the other hand, it was shown that the sequential method ME2 can fail to converge when the links which do not have flows loaded in the first all or nothing assignment are observed for matrix estimation. In this section, this topic will be further investigated numerically using the Nguyen's example network which is more complicated than that analyzed earlier.

A test was designed to investigate the effects on the convergence of the estimation methods due to the location of observations of link flows. This can be carried out by comparing the results from the tests using three different sets of observed link flows. The first test is to estimate a trip matrix from flows observed on the links which are loaded in the first all or nothing assignment. The second one is to estimate a trip matrix from flows observed on links which are not loaded in the first all or nothing assignment. The third one is to estimate a trip matrix from link flows observed for a mixed selection of links. An observed trip matrix (300,300,300,300) was assigned to obtain the observed link flows. The selection of links for each of three tests was made by examining the assigned flows from the first all or nothing

assignment. Among the links - (12,8), (6,7), (6,10), (9,10) and (9,13) - located on a cordon line, the links (6,7) and (9,10) have flows loaded in the first all or nothing assignment but the links (12,8), (6,10) and (9,13) do not. Accordingly, the first test was designed to estimate a trip matrix from link flows observed on the links (6,7) and (9,10). The second test was designed to estimate a trip matrix from link flows observed on the links (12,8), (6,10) and (9,13). The third test was designed to estimate a trip matrix from link flows observed on the links (6,7) and (6,10). A trip matrix (350,250,150,50) which is not uniformly scaled from the observed one was used as a prior trip matrix in all these tests.

6.3.8.2 Results

Figure 6.12a shows the gap achieved by the sequential ME2 method during the first 20 iterations. It can be seen that this method reduces the gap rapidly in the first test in which only links with flows in the first all or nothing assignment are observed for matrix estimation. However, the sequential ME2 method fails to reduce the gap in the second and third tests in which links without flows in the first all or nothing assignment are also observed for matrix estimation. These results contrast with the results obtained by the simultaneous methods SIM-S₁ and SIM-S₀ as shown in Figures 6.12c and 6.12e.

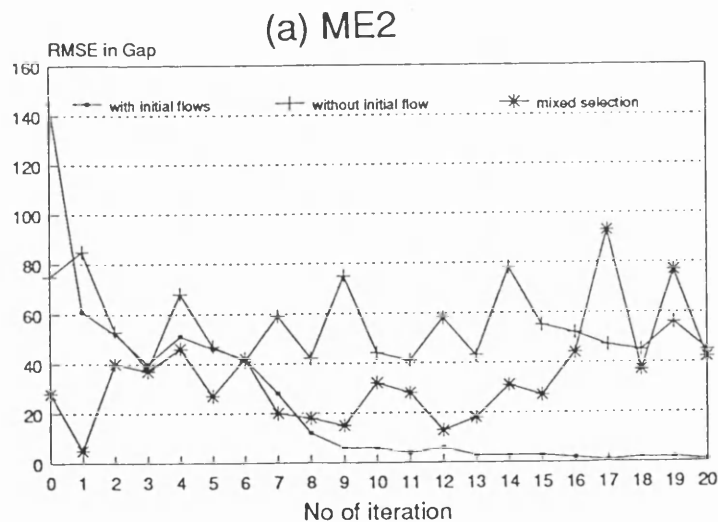
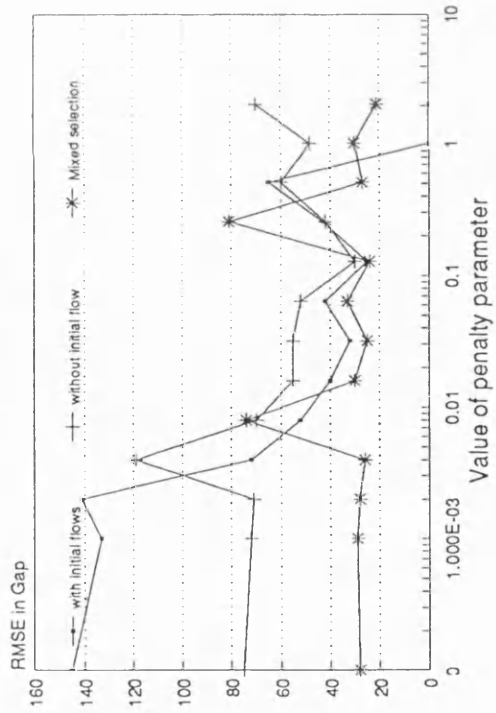
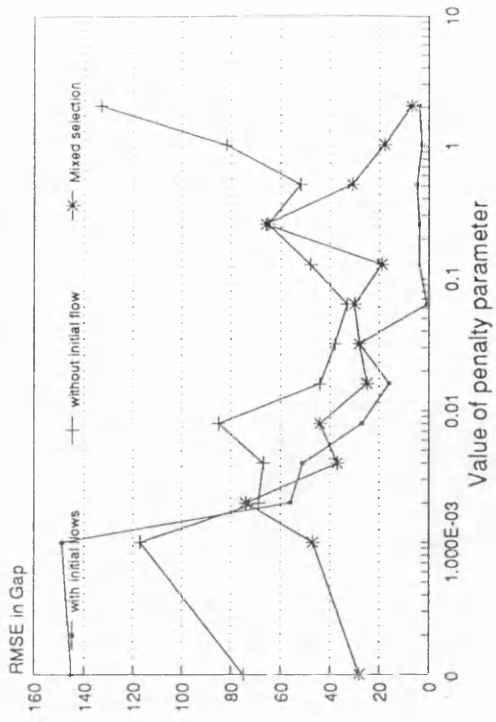


Figure 6.12 Results of the tests with different sets of links observed

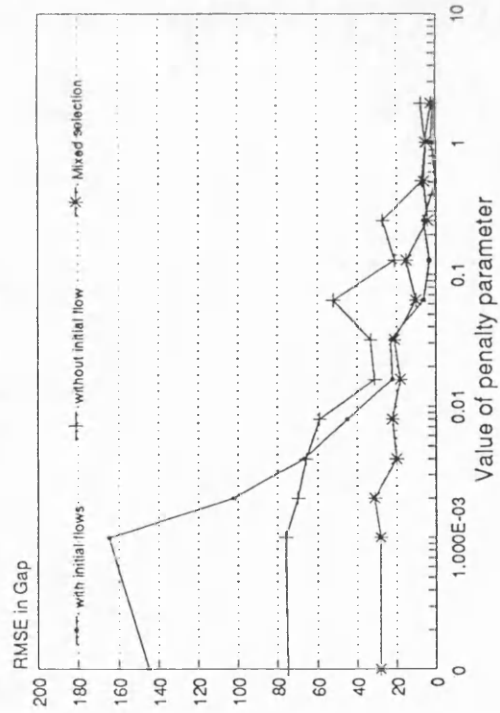
(b) ME3-S1



(d) ME3-So



(c) SIM-S1



(e) SIM-So

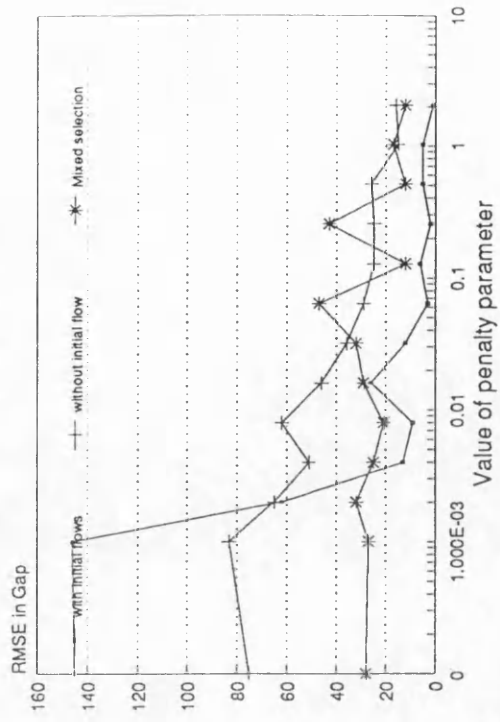


Figure 6.12 Results of the tests with different sets of links observed

Inspection of the results of the simultaneous methods indicates that it reduces the gap well in all the three tests. Thus, the simultaneous methods are less dependent upon the the selection of the observed links for matrix estimation. Figures 6.12b and 6.12d show the result by the improved sequential methods ME3-S1 and ME3-S0. Both of the improved sequential methods ME3-S1 and ME3-S0 have severe oscillations in the second and third tests in which links without initial flows are observed for matrix estimation. As shown in Figures 6.12c and 6.12e, the simultaneous methods SIM-S1 and SIM-S0 produce similar results to each other. This indicates that the use of different entropy functions makes no substantial difference on the performance of the estimation methods. In the case of the improved sequential methods, as shown in Figure 6.12b and 6.12d, use of the entropy function $S_0(\underline{T}, t)$ produces even worse results compared with those of $S_1(\underline{T}, t)$.

The results of the tests by all of the estimation methods are summarized in Tables 6.12a-c. It can be said from these results that the simultaneous method performs best in terms of entropy values and gap values, thus indicating their robustness with respect to selection of links for flow observations.

Table 6.12 Results of the tests with different sets of links observed

(a) Using only links with initial flows

Solution Methods	Trip Matrices					Entropy		Gap		Fit	
	T ₁₃	T ₁₄	T ₂₃	T ₂₄	T _{..}	S ₀	S ₁	RMSE	%RMSE	RMSE	%RMSE
Observed	300	300	300	300	1200	-267	446	0	0	0	0
Prior	350	250	150	50	800	0	800	145	63	150	50
ME 2	126	90	418	353	986	-691	89	1	0	151	50
ME 3 - S ₁	84	58	407	384	933	-842	-52	1	0	176	59
SIM - S ₁	469	161	416	107	1153	-151	581	2	0	157	52
ME 3 - S ₀	538	389	447	106	1480	-61	509	0	0	176	59
SIM - S ₀	541	453	432	87	1513	-46	503	5	1	190	63

Table 6.12 (cont.)

(b) Using only links without initial flows

Solution Methods	Trip		Matrices			Entropy		Gap		Fit	
	T ₁₃	T ₁₄	T ₂₃	T ₂₄	T _{..}	S ₀	S ₁	RMSE	%RMSE	RMSE	%RMSE
Observed	300	300	300	300	1200	-267	446	0	0	0	0
Prior	350	250	150	50	800	0	800	7.5	44	150	50
ME2	514	171	150	386	1221	-405	300	45	27	152	51
ME3 - S ₁	309	183	150	334	976	-344	438	39	23	97	32
SIM - S ₁	376	246	251	262	1135	-188	550	26	16	56	19
ME3 - S ₀	467	247	233	302	1249	-221	471	55	33	94	31
SIM - S ₀	328	261	259	320	1168	-283	443	2	1	33	11

(c) Using links both with or without initial flows

Solution Methods	Trip		Matrices			Entropy		Gap		Fit	
	T ₁₃	T ₁₄	T ₂₃	T ₂₄	T _{..}	S ₀	S ₁	RMSE	%RMSE	RMSE	%RMSE
Observed	300	300	300	300	1200	-267	446	0	0	0	0
Prior	350	250	150	50	800	0	800	28	14	150	50
ME2	361	73	374	50	858	-203	595	42	21	176	59
ME3 - S ₁	455	263	417	50	1185	-93	626	5	2	159	53
SIM - S ₁	309	278	289	46	922	-46	745	2	1	128	43
ME3 - S ₀	388	217	167	51	824	-5	795	34	17	153	51
SIM - S ₀	328	276	283	55	941	-37	751	1	0	124	41

6.3.8.3 Conclusions

It can be said from these results that the sequential ME2 method and the improved sequential methods ME3-S₁ and ME3-S₀ fail to converge when flows are observed for matrix estimation on links which do not have the initial flows in the first all or nothing assignment. On the other hand, the simultaneous methods SIM-S₁ and SIM-S₀ perform well with respect to selection of links for matrix estimation.

From these results, we might make a useful suggestion for link observations when estimating trip matrices from traffic counts. That is, the observation of links which have the initial flows in the first all or nothing assignment could help the sequential method or the improved sequential method to converge.

6.4 Tests with the Ripon network

6.4.1 Introduction

In the previous sections, small example networks have been used to test the proposed estimation methods. These tests were useful, because a variety of different cases could be investigated effectively using synthetic data generated artificially within a limited amount of computing time. However, the tests using small artificial networks can not tell how likely the method is to perform well in practice. In order to do this, it is necessary to test the proposed estimation methods using realistic networks. The test of an estimation method for estimating a trip matrix from traffic counts requires a real data set ideally consisting of:

- (1) network data e.g. link distance, capacity, speed-flow relationship,
- (2) an independently observed trip matrix of reasonable accuracy, obtained for example through direct survey methods.
- (3) a set of observed link flows on the network,

This type of data set, in particular its second element, is difficult to obtain.

A data set collected in the town of Ripon was examined for an application to this study. Ripon is a busy market town in North Yorkshire, England, lying north of Harrogate on the A61. North Yorkshire County Council (NYCC) was considering plans for a by-pass for Ripon and therefore conducted O-D surveys in May 1978 and May 1985. These two surveys, in conjunction with separate traffic counts collected by Steer, Davies and Gleave Ltd (SDG) in November 1985, form a basis of this data set (Steer, Davies and Gleave, 1987). The Ripon data base consists of the following data sets:

- (1) **The 1985 network description:** A map of the town of Ripon is depicted in Figure 6.13 and the 1985 network description is shown in Figure 6.14.

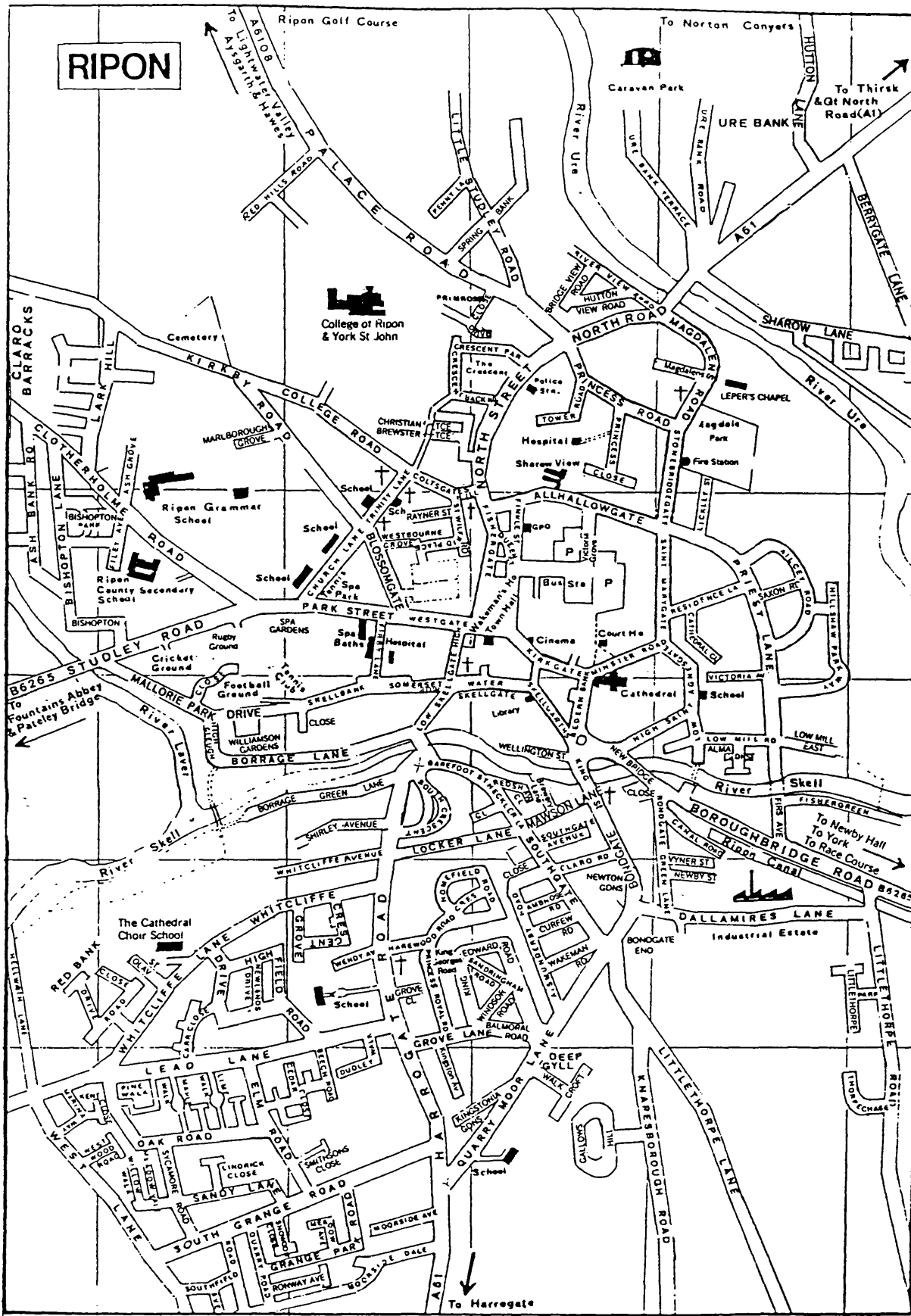


Figure 6.13 Map of the town Ripon

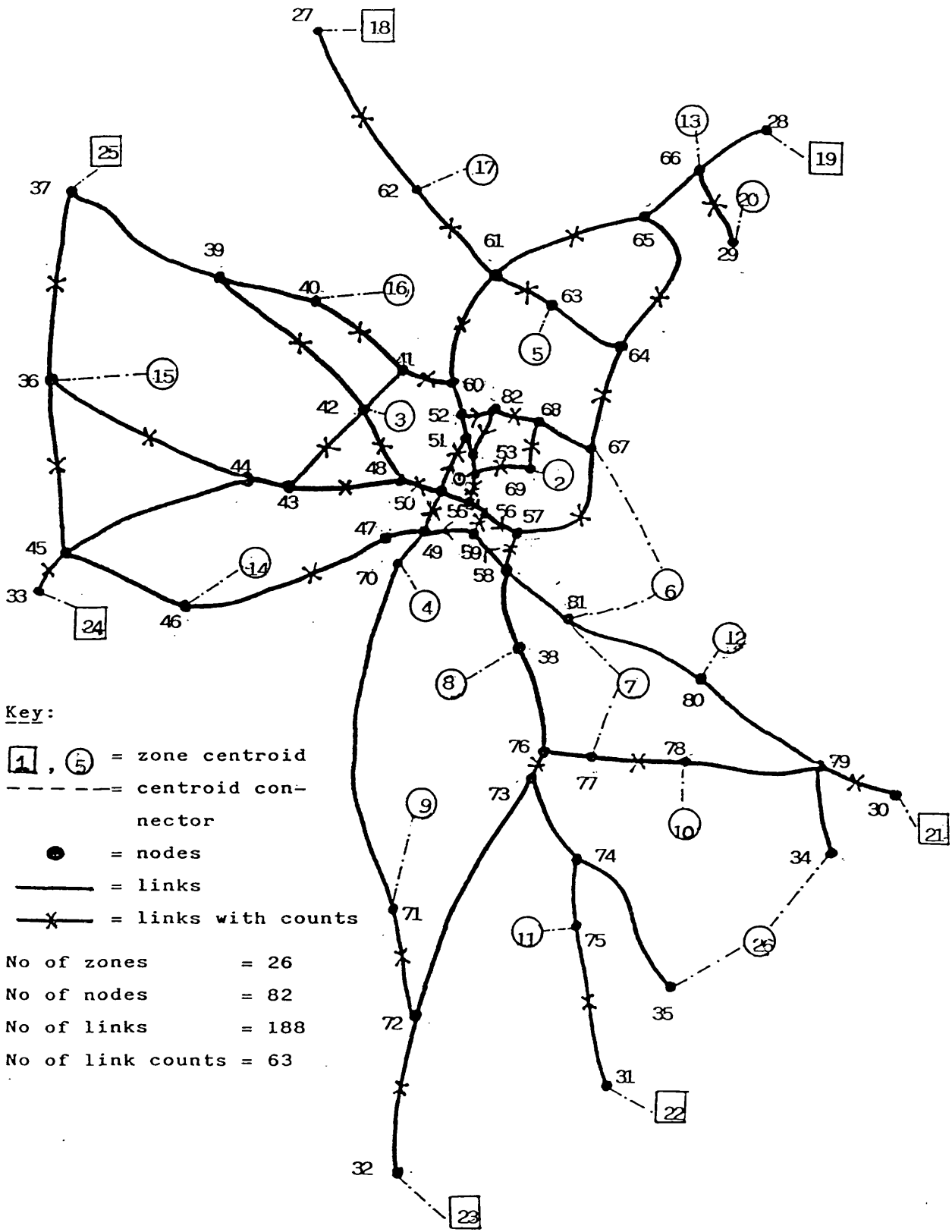


Figure 6.14 Ripon network

The study area is divided into 26 zones of which 19 are internal and 7 are external. The road network contains 82 nodes and 188 one-way links. In order to overcome congestion around the city centre during the peak hour, a one-way scheme was introduced and this was also modelled.

(2) **The 1985 observed trip matrix:** This is a 24 hour, annual average daily traffic (AADT) trip matrix, compiled from a roadside interview survey collected by NYCC in May 1985. The 1985 observed trip matrix was used as the basis of the observed trip matrix for the test.

(3) **The 1985 observed traffic counts:** There are 63 observed traffic counts collected throughout Ripon by Steer, Davies and Gleave (SDG) Ltd in November 1985 and by NYCC in May 1985 in conjunction with their roadside interview survey. The sites of collecting traffic counts are also shown in Figure 6.14. Two automatic traffic counters were laid down by NYCC. One was on the main North-South route (between nodes 60-61) and the other was on the main West-east route (between nodes 43-48) through Ripon. All other counts were taken manually. The automatic traffic counts enabled SDG to estimate a factor to convert the results of manual counts to 24 hour counts. A factor was derived for each site, in each direction, for each day of the counts. As eight factors only ranged from 1.7 to 1.8, the uniform factor of the flow-weighted mean value 1.758 was applied to the manual traffic counts. The traffic counts were found to be similar between May and November 1985 given that some seasonal variation is to be expected and that the November traffic counts were taken over one day only. All of the November traffic counts are within $\pm 10\%$ of the May counts at the sites where a comparison is possible.

(d) **The prior trip matrix:** Besides the data sets mentioned above, the matrix estimation method may require a prior trip matrix. The 1978 trip matrix compiled from roadside interview data by NYCC could serve as a good prior. However, this data was not available for the present study. As an alternative, in this study a uniform prior trip matrix which is considered to be close to the 1985 observed trip matrix was used. The uniform prior trip matrix used in this test was created by minimising the gap between its assigned link flows and the link flows assigned from the 1985 observed one.

The test in this study requires a peak hour real trip matrix and peak hour traffic counts corresponding to the real trip matrix for the test. However, the 1985 observed trip matrix and 1985 traffic counts were only available as whole day data. It is not possible to convert 24 hour trip matrix and 24 hour traffic counts to peak hour trip matrix and traffic counts by using peak hour scaling factor such as those given in Traffic Appraisal Manual (DTp, 1981). The use of a peak hour factor gives only peak hour data uniformly scaled from whole day data. The real peak hour trip matrix might be quite different from these uniformly scaled data for example due to the existence of tidal flows between O-D pairs. Furthermore, the scaled peak hour real trip matrix has no corresponding relationship with peak hour traffic counts because of the inherent non-linearity of the assignment process. Consequently, it is not possible to perform an ideal test by using only the data sets available in the Ripon data base.

Under these constraints, the following two tests were designed to investigate the performance of the estimation methods. The first test was to estimate a trip matrix from artificial traffic counts. These counts were the modelled flows obtained by assigning the peak hour trip matrix to the Ripon road network. The peak trip matrix used here was obtained from the 1985 AADT real trip matrix by using peak hour factor $2.630/24 \cong 0.10958$ given in Table 5A of Appendix D14 in Traffic Appraisal Manual (DTp, 1981). Therefore, traffic counts used in the first test are synthetic and error-free. The main reason to carry out this artificial test is to investigate the performance of the various matrix estimation methods in a realistic network. The second test was to estimate a trip matrix from peak hour traffic counts which is not error-free. Peak hour traffic counts used in the second test were obtained from the 1985 AADT real traffic counts by the same procedure used to calculate the peak hour observed trip matrix. In the second test, however, the comparison between the estimated trip matrix and the observed one is not possible, as the real trip matrix corresponding to traffic counts is not available. This second test was intended to investigate how well the various estimation methods accommodate inconsistent traffic counts in a real network.

The SATURN program (Van Vliet, 1987) which implements the ME2 method and whose assignment program is interfaced with the program of the proposed estimation methods was used to code the network data. As the computing tool, the mainframe AMDAHL 5890 at the University of London Computing Centre (ULCC) was used for the sequential method ME2 and the improved sequential methods ME3-S₁ and ME3-S₀. The super-computer CRAY X-MP at ULCC was used for the simultaneous methods SIM-S₁ and SIM-S₀.

6.4.2 Tests with assigned link flows

6.4.2.1 Introduction

It is interesting to see how accurately the estimation methods can reproduce the original trip matrix from the error-free synthetic traffic counts. By assigning the original trip matrix onto the Ripon network using equilibrium assignment, we obtain the assigned flows on links which contain no sample or route choice errors. A subset of these flows was selected for the links as shown in Figure 6.14. These selected link flows were used with each of the various methods to estimate a trip matrix. As mentioned above, a uniform prior trip matrix was used with cell values t_{ij} each equal to 5.

6.4.2.2 Results

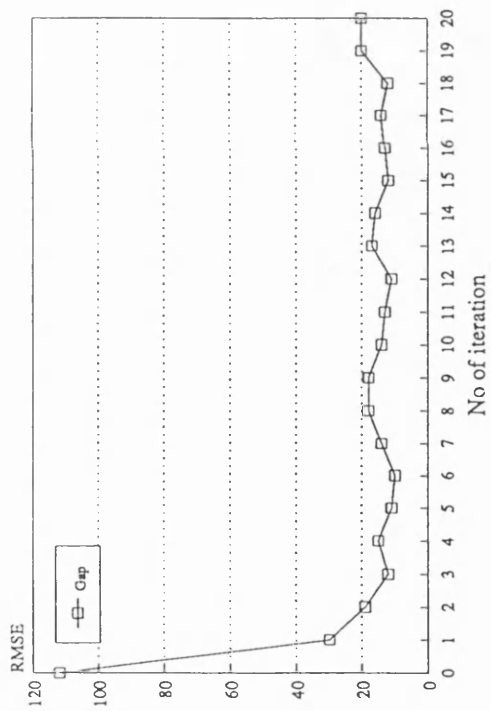
The performance of the sequential estimation method of ME2 can be investigated by examining indicators such as totals of estimated trip matrices, entropy values, gap values and goodness of fit, as the number of iterations between equilibrium assignment and matrix estimation increases. The results obtained by ME2 are summarized in Table 6.13 and presented graphically in Figures 6.15a-d. Figure 6.15a shows the evolution of the gap. It shows that the sequential method reduces gap rapidly during the first two iterations and makes relatively small oscillations to it thereafter. In particular, most of reduction in gap is made during the first iteration. This suggests that congestion has little effects on route choice proportions in the Ripon network and the use of fixed values for them is not bad. A similar result is obtained for the totals of estimated trip matrices, as shown in Figure 6.15c.

Table 6.13 Results of the test using assigned link flows by ME2

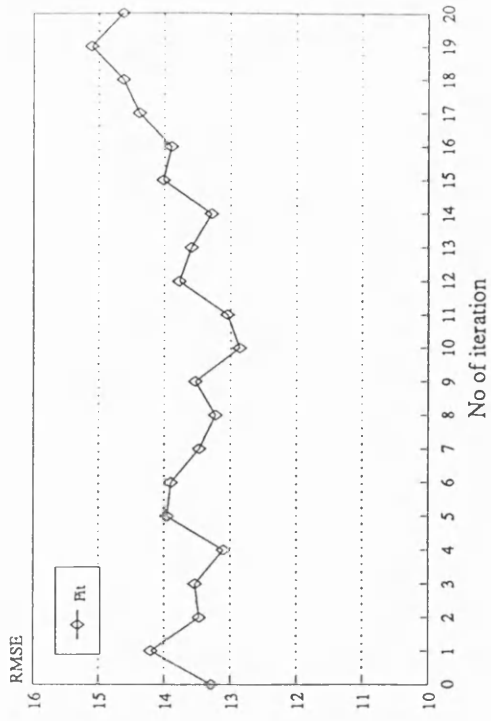
Iteration	T..	Entropy		Gap		Fit	
		S0	S1	RMSE	%RMSE	RMSE	%RMSE
Observed	3949	-4111	-67	0	0	0	0
Prior	3250	0	3250	112	48	13.29	219
1	4552	-2423	596	30	12	14.20	234
2	4417	-2334	728	19	8	13.47	222
3	4521	-2347	683	12	5	13.53	223
4	4533	-2333	692	15	6	13.12	216
5	4638	-2481	508	11	4	13.96	230
6	4658	-2459	523	10	4	13.90	229
7	4622	-2462	532	14	5	13.47	222
8	4558	-2399	618	18	7	13.23	218
9	4620	-2529	466	18	7	13.53	223
10	4517	-2176	855	14	5	13.87	212
11	4568	-2315	698	13	5	13.05	215
12	4606	-2578	422	11	4	13.78	227
13	4582	-2427	581	17	7	13.52	224
14	4523	-2386	642	16	7	13.22	219
15	4679	-2733	241	12	5	14.02	231
16	4652	-2593	391	13	5	13.90	229
17	4684	-2869	103	14	6	14.38	237
18	4668	-3057	-79	12	5	14.63	241
19	4693	-3186	-217	20	8	15.11	249
20	4626	-3139	-146	20	8	14.63	241

It shows that ME2 estimates the total of about 4600 trips after the first iteration and makes small changes to it thereafter. However, as shown in Figure 6.15b, an inspection of the goodness of fit between the estimated trip matrices and the observed trip matrix shows that they are subject to considerable fluctuations throughout the 20 iterations of the test. It shows that the sequential ME2 method fails to improve the goodness of fit compared with that of the prior trip matrix. Figure 6.15d shows the trade-off relationship between entropy (S1) values and gap values of estimated trip matrices. It shows that during the first two iterations substantial improvements are made in both gap and entropy, but it is difficult to identify any particular relationship thereafter.

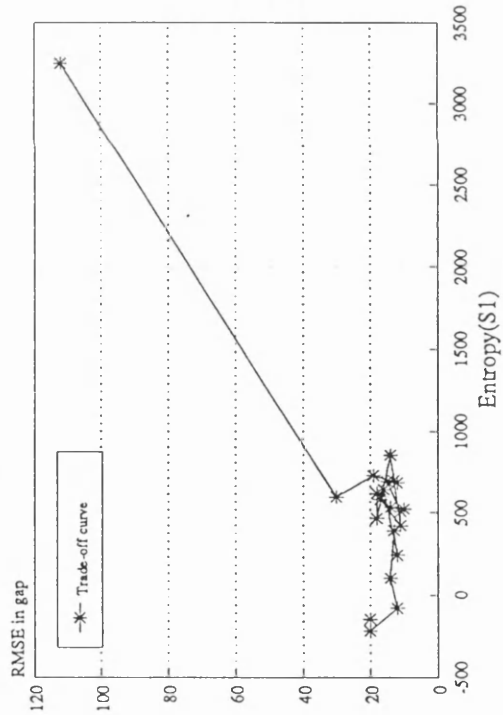
(a) Evolution of gap



(b) Evolution of goodness of fit



(d) Trade-off curve



(c) Evolution of total demand

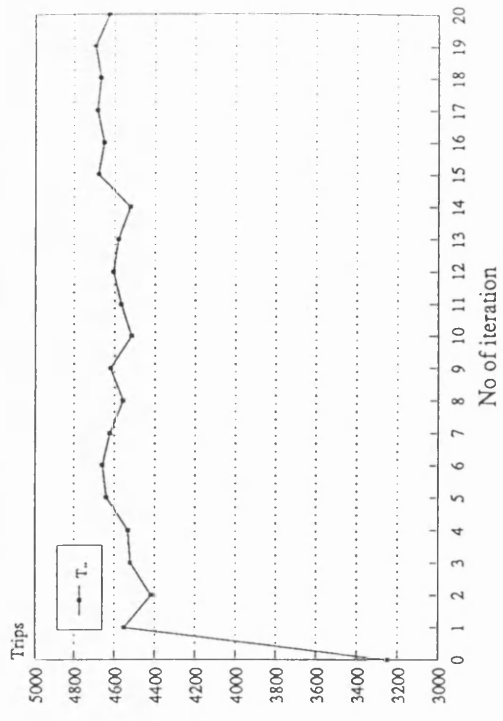


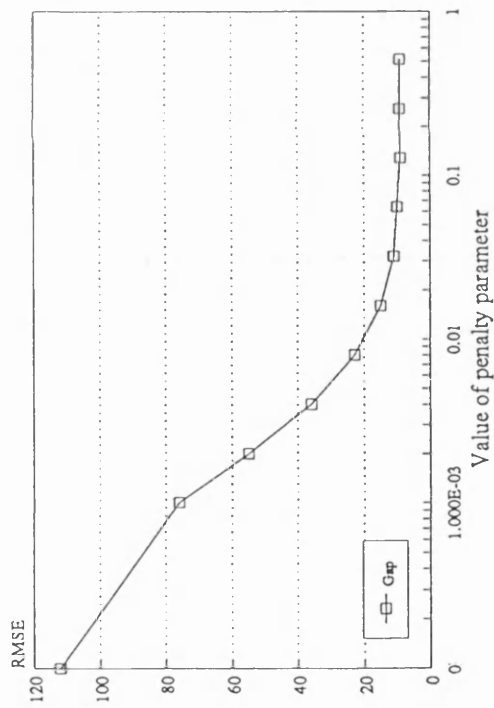
Figure 6.15 Results of the test using assigned link flows by ME2

The simultaneous method SIM-S₁ estimates a trip matrix by using a sequential unconstrained optimization method which approximates the solution of the original estimation problem with increasing accuracy by progressively increasing the magnitude of the penalty parameter. The results obtained by SIM-S₁ are summarized in Table 6.14 and presented graphically in Figures 6.16a-d. Figure 6.16a shows reduction in gap as the magnitude of the penalty parameter increases. It shows that SIM-S₁ reduces the gap value gradually during the first six subproblems of the sequence and it makes small reductions to it thereafter. Figure 6.16b shows the evolution of the goodness of fit of the estimated trip matrices. It shows that SIM-S₁ improves the goodness of fit during the first six subproblems of the sequence and thereafter the goodness of fit becomes worse. Figure 6.16c shows the totals of estimated trip matrices. It shows that SIM-S₁ estimates the total of about 4600 trips after the first five subproblems. Figure 6.16d shows the trade-off relationship between gap and entropy values. It shows that a clear trade-off curve exists. It is noted that after the first six subproblems the entropy value is further reduced without reducing the gap. A cross-examination of Figures 6.16b and 6.16d indicates that SIM-S₁ improves the goodness of fit during the first five sub-problems in which major reduction in gap is made.

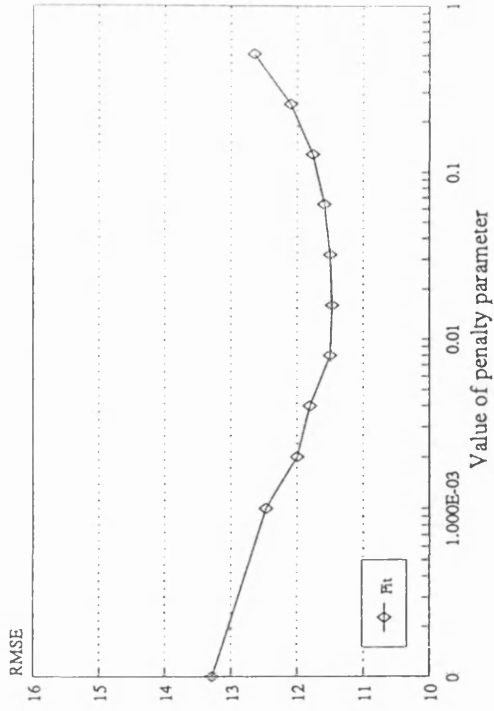
Table 6.14 Results of the test using assigned link flows by SIM-S₁

Penalty Parameter	T..	Entropy		Gap		Fit	
		S0	S1	RMSE	%RMSE	RMSE	%RMSE
0.000	3250	0	3250	112	45	13.29	219
0.001	4103	-134	2772	76	30	12.47	205
0.002	4320	-319	2443	55	22	12.00	198
0.004	4503	-592	2115	36	14	11.81	195
0.008	4612	-883	1839	23	9	11.51	190
0.016	4693	-1130	1599	15	6	11.48	189
0.032	4759	-1345	1314	11	4	11.51	190
0.064	4844	-1597	1050	10	4	11.60	191
0.128	4921	-1829	688	9	4	11.77	194
0.256	4967	-2172	223	9	4	12.10	199
0.512	5123	-2569	223	9	4	12.65	208

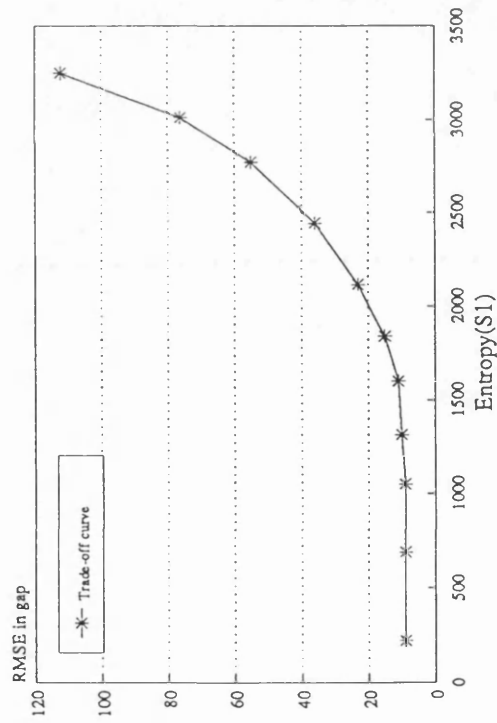
(a) Evolution of gap



(b) Evolution of goodness of fit



(d) Trade-off curve



(c) Evolution of total demand

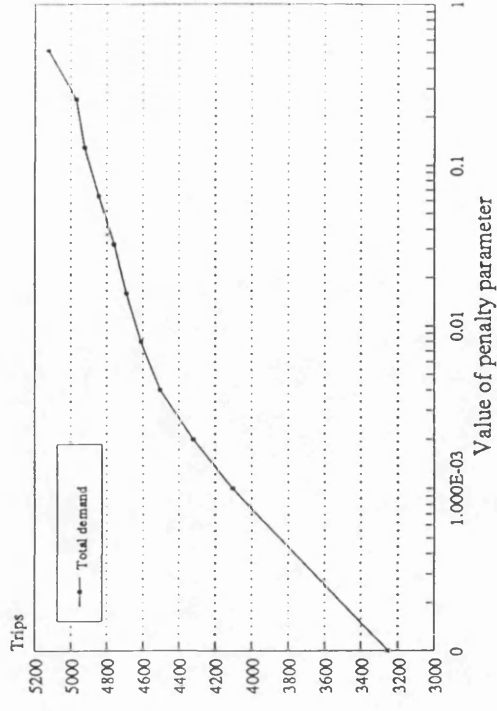


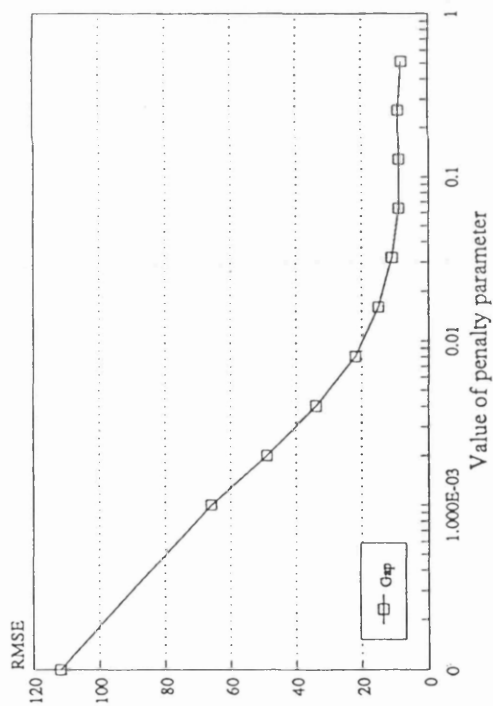
Figure 6.16 Results of the test using assigned link flows by SIM-S1

A similar result to SIM-S₁ was made by SIM-S₀. This result is shown in Table 6.15 and depicted graphically in Figures 6.17a-d. As shown in Figure 6.17c, a main difference in the performance between the methods SIM-S₁ and SIM-S₀ is that the total demands of the trip matrices estimated by SIM-S₀ are higher than those by SIM-S₁, specially during the early subproblems.

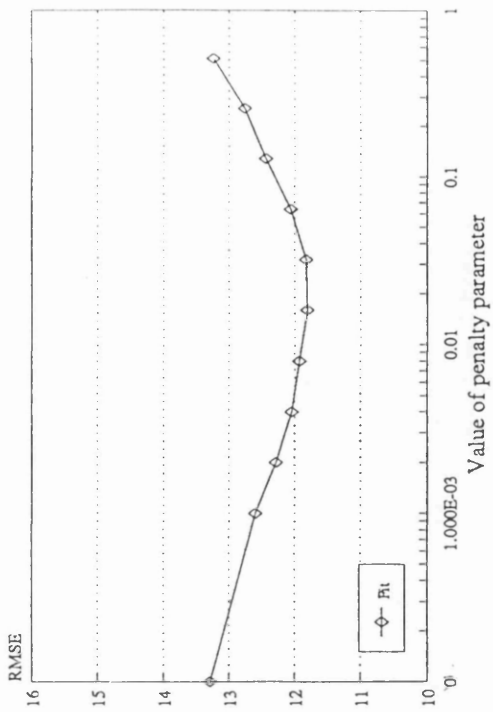
Table 6.15 Results of the test using assigned link flows by SIM-S₀

Penalty Parameter	T..	Entropy		Gap		Fit	
		S0	S1	RMSE	%RMSE	RMSE	%RMSE
0.000	3250	0	3250	112	45	13.29	219
0.001	4916	-131	2750	66	26	12.59	208
0.002	4985	-319	2533	49	20	12.29	202
0.004	5019	-570	2267	34	14	12.05	199
0.008	5069	-869	1946	22	9	11.94	197
0.016	5085	-1115	1694	15	6	11.81	195
0.032	5115	-1340	1455	11	4	11.83	195
0.064	5216	-1603	1145	9	4	12.06	199
0.128	5252	-1976	755	9	4	12.44	205
0.256	5356	-2225	455	9	4	12.76	210
0.512	5381	-2661	7	8	3	13.24	218

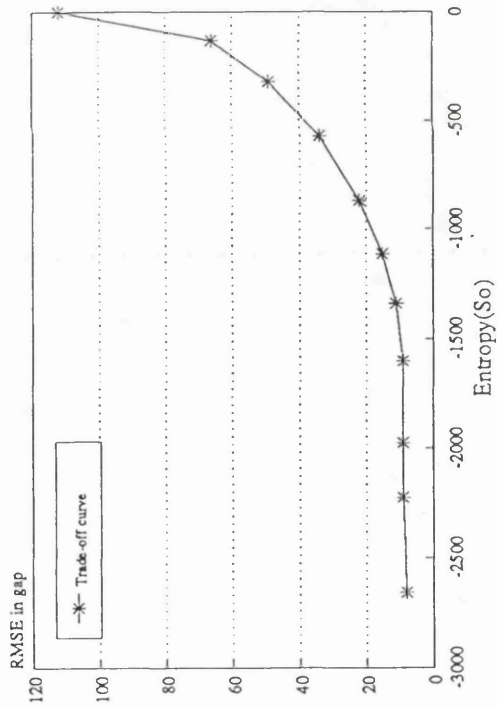
(a) Evolution of gap



(b) Evolution of goodness of fit



(d) Trade-off curve



(c) Evolution of total demand

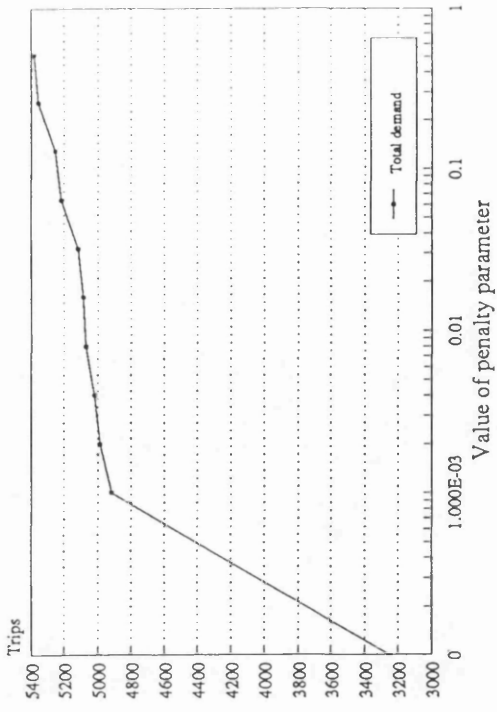


Figure 6.17 Results of the test using assigned link flows by SIM-S₀

Further similar results to the ones by the simultaneous methods were produced by the improved sequential methods ME3-S₀ and ME3-S₁. The results of the improved sequential methods are summarized in Tables 6.16 and 6.17. The results in Tables 6.16 and 6.17 are presented graphically in Figures 6.18a-d and Figures 6.19a-d.

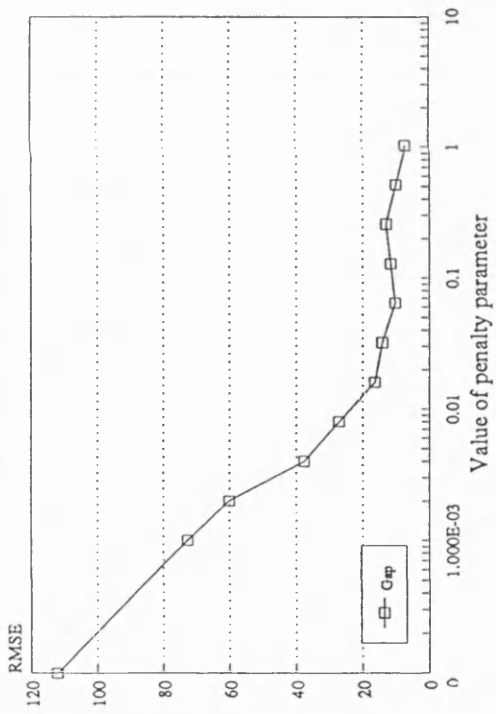
Table 6.16 Results of the test using assigned link flows by ME3-S₁

Penalty Parameter	T..	Entropy		Gap		Fit	
		S0	S1	RMSE	%RMSE	RMSE	%RMSE
0.000	3250	0	3250	112	45	13.29	219
0.001	3821	-114	3088	73	29	12.52	206
0.002	4030	-313	2850	60	24	12.17	200
0.004	4166	-627	2504	38	15	11.73	193
0.008	4276	-922	2180	27	11	11.55	190
0.016	4309	-1211	1883	17	7	11.72	193
0.032	4378	-1411	1662	14	6	11.65	192
0.064	4351	-1644	1438	10	4	11.70	193
0.128	4468	-1869	1177	12	5	11.99	197
0.256	4480	-1988	1054	13	5	12.09	199
0.512	4447	-2120	933	10	4	12.23	201

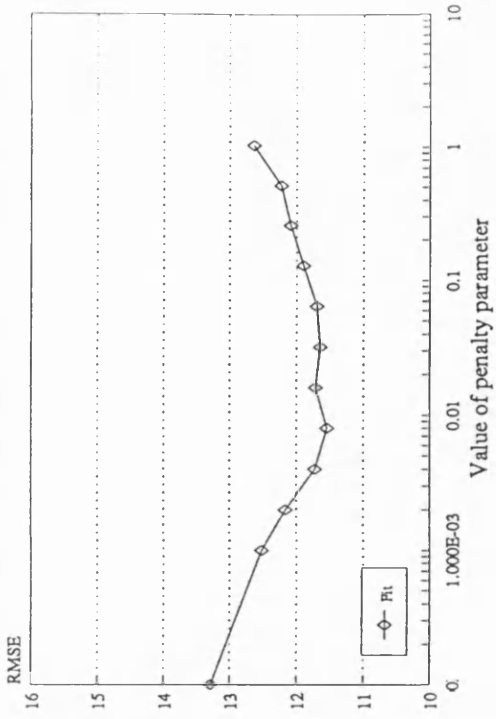
Table 6.17 Results of the test using assigned link flows by ME3-S₀

Penalty Parameter	T..	Entropy		Gap		Fit	
		S0	S1	RMSE	%RMSE	RMSE	%RMSE
0.000	3250	0	3250	112	45	13.29	219
0.001	4525	-114	2914	70	28	12.58	207
0.002	4441	-301	2754	55	22	12.21	201
0.004	4532	-548	2478	38	15	11.86	195
0.008	4634	-897	2093	30	12	11.82	195
0.016	4514	-1197	1834	16	6	11.65	192
0.032	4640	-1468	1520	13	5	11.87	195
0.064	4633	-1601	1389	14	6	11.82	195
0.128	4679	-1943	1031	15	6	12.23	201
0.256	4694	-1965	1003	17	7	12.27	202
0.512	4773	-2325	613	13	5	12.81	211

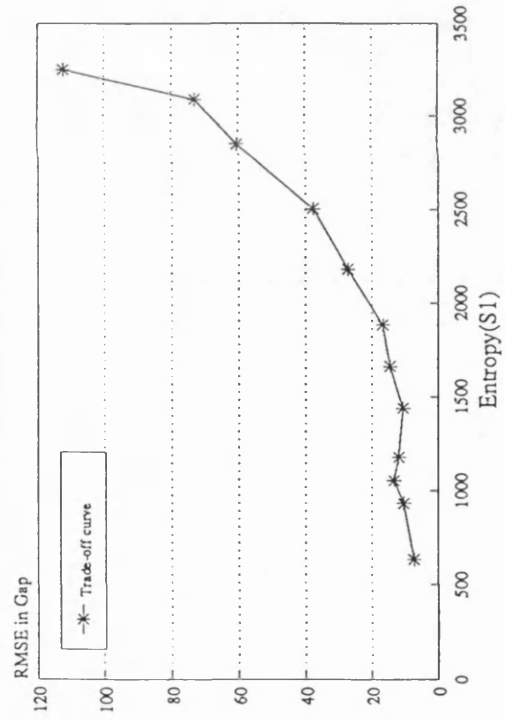
(a) Evolution of gap



(b) Evolution of goodness of fit



(d) Trade-off curve



(c) Evolution of total demand

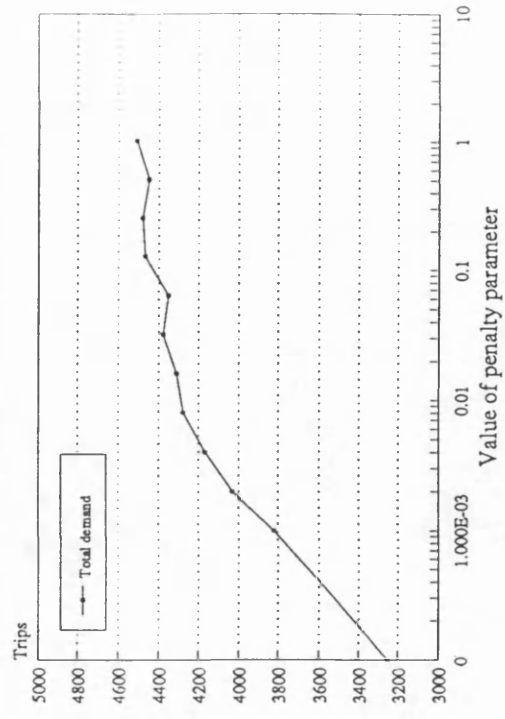
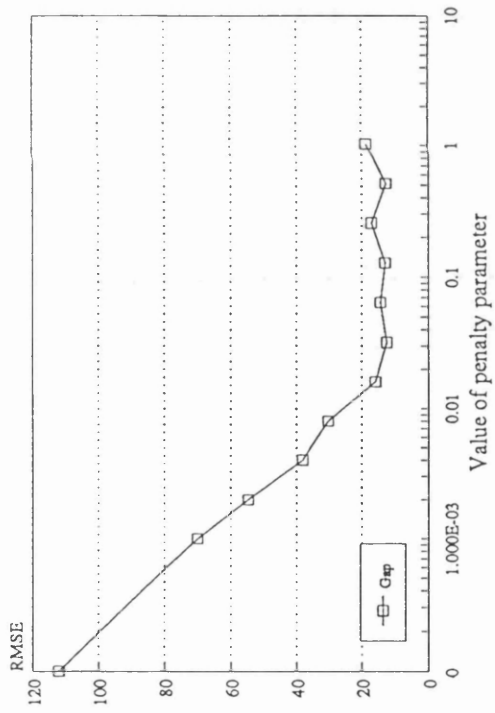
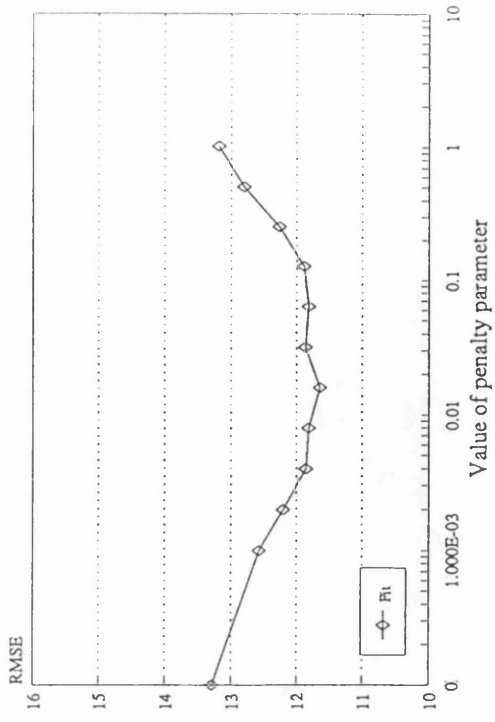


Figure 6.18 Results of the test using assigned link flows by ME3-S1

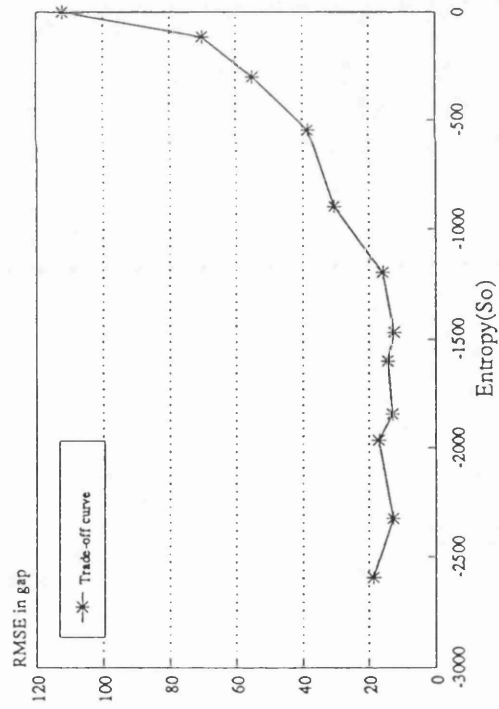
(a) Evolution of gap



(b) Evolution of goodness of fit



(d) Trade-off curve



(c) Evolution of total demand

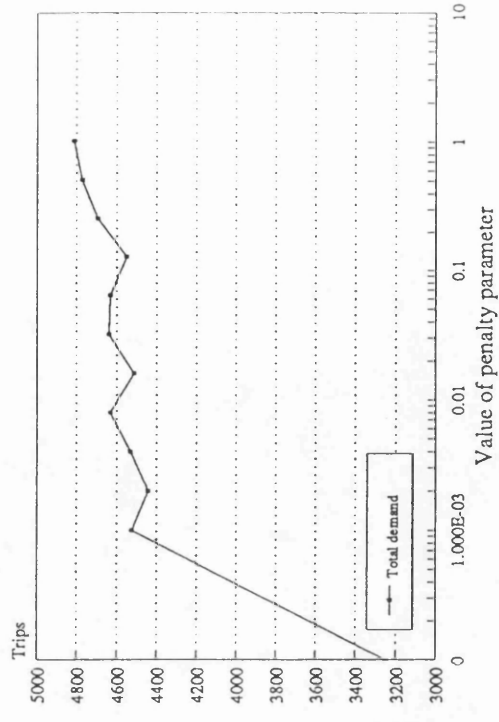


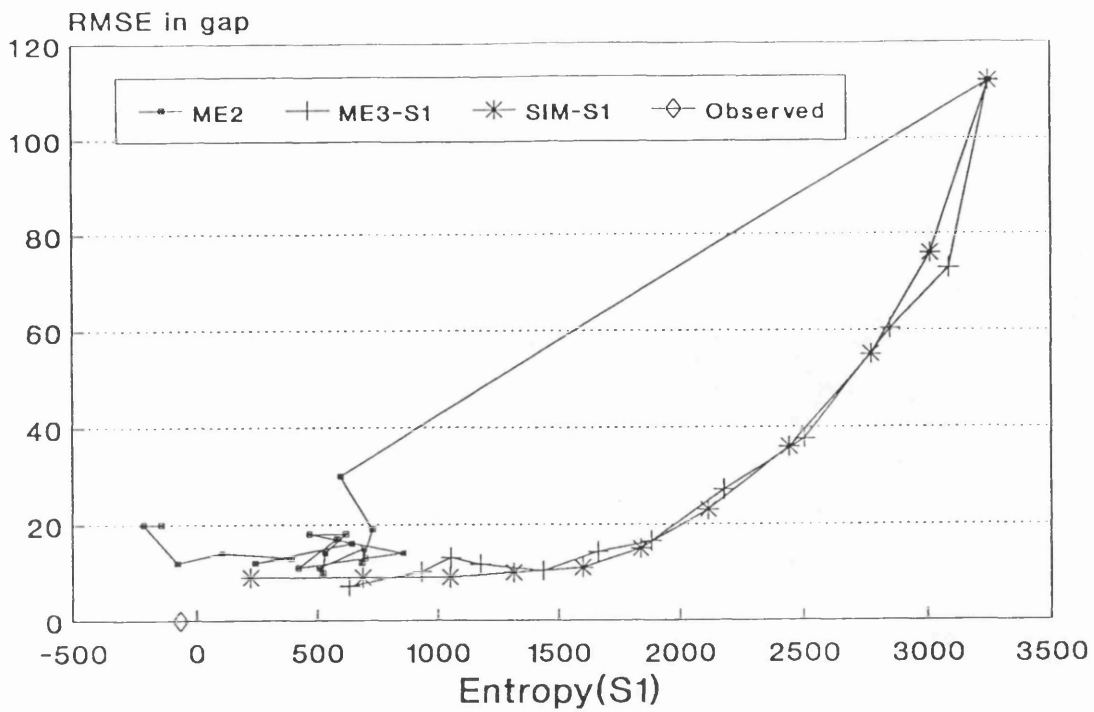
Figure 6.19 Results of the test using assigned link flows by ME3-So

Figures 6.18a and 6.19a show the evolution of the gap value, as the magnitude of the penalty parameter increases. ME3-S₀ and ME3-S₁ reduce the gap value gradually during the first five subproblems and thereafter they make small changes to it. Figures 6.18b and 6.19b show the goodness of fit between the estimated trip matrices and the observed one. They show that ME3-S₀ improves the goodness of fit during the first five subproblems and thereafter the goodness of fit becomes worse. Similarly, ME3-S₁ improves the goodness of fit during the first four subproblems and thereafter the goodness of fit becomes worse. Figures 6.18c and 6.19c show the totals of the estimated trip matrices for each method. They show that ME3-S₀ estimates the total of about 4500 trips during the first subproblem and thereafter it makes relatively small changes. On the other hand, ME3-S₁ approaches the total of about 4500 trips gradually through the sequence. A comparison of the totals estimated by these methods suggests that ME3-S₀ tends to overestimate and ME3-S₁ tends to underestimate. Figures 6.18d and 6.19d show trade-off curves between gap and entropy values. They show that a clear trade-off curve exists in each case.

For comparison purposes, the trade-off curves of three alternative estimation methods ME2, ME3-S₁ and SIM-S₁ are all shown together in Figure 6.20a and those of ME3-S₀ and SIM-S₀ are also shown in Figure 6.20b. Figure 6.20a shows that ME3-S₁ and SIM-S₁ perform similarly and they perform better than ME2 in terms of entropy and gap values. Also, Figure 6.20b shows that ME3-S₀ and SIM-S₀ perform similarly, although SIM-S₀ produces slightly better performance than ME3-S₁.

Another comparison is made for the goodness of fit of the estimated trip matrices by all five estimation methods ME2, ME3-S₁, ME3-S₀, SIM-S₁ and SIM-S₀, as shown in Figure 6.21. Figure 6.21 shows that ME3-S₁, ME3-S₀, SIM-S₁ and SIM-S₀ all perform similarly and their goodness of fit of estimated trip matrices are all better than that of ME2.

(a) By the methods using the entropy function $S_1(T,t)$



(b) By the methods using the entropy function $S_0(T,t)$

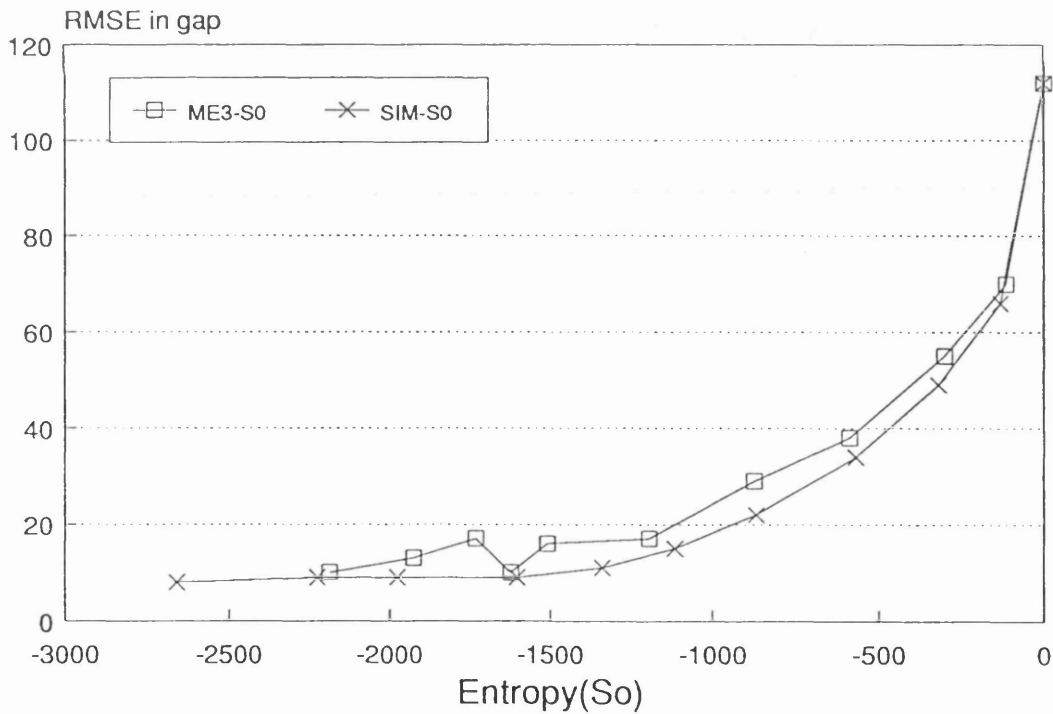


Figure 6.20 Comparison of the trade-off curves estimated from assigned link flows by various solution methods

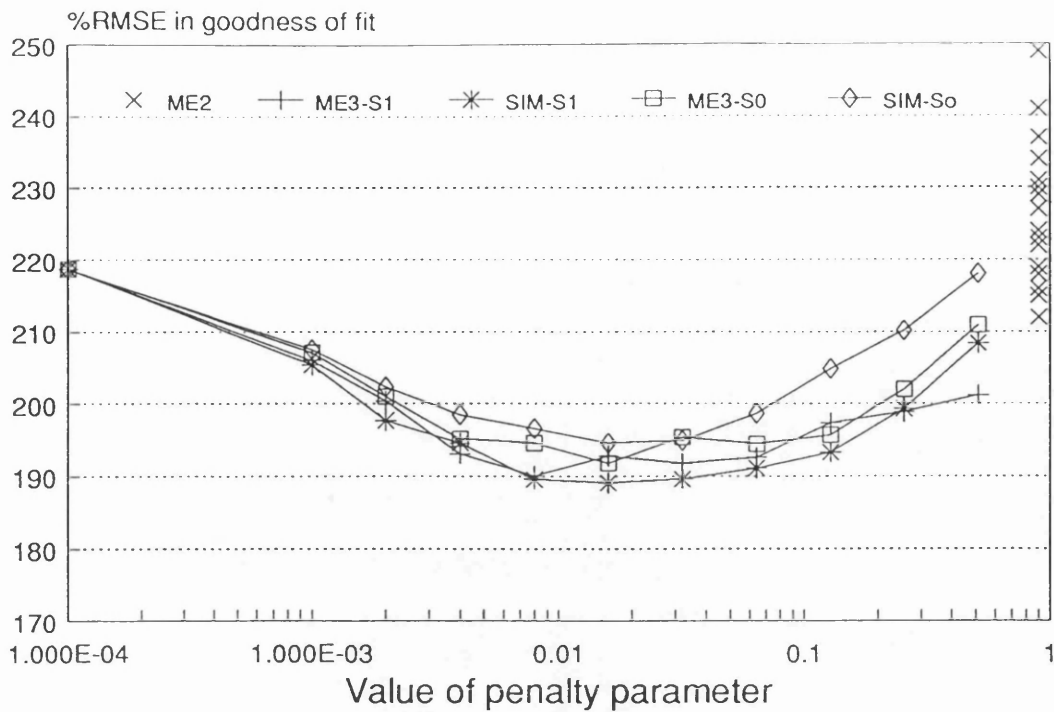
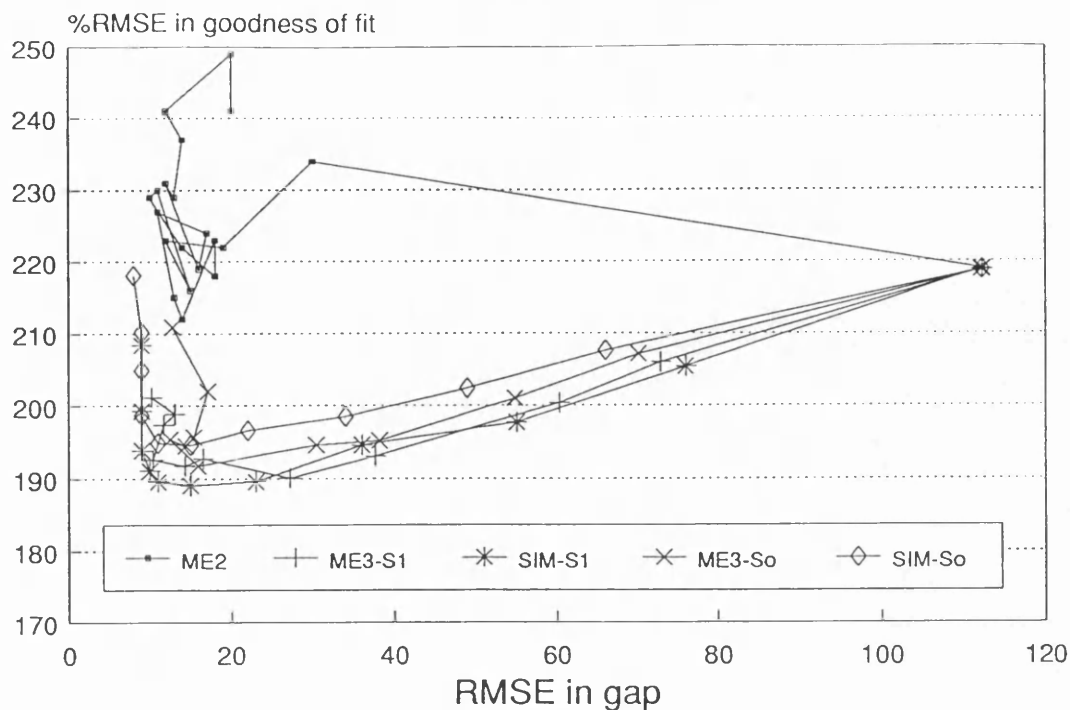


Figure 6.21 Comparison of the goodness of fit of the trip matrices estimated by various solution methods

Further comparisons are made for the goodness of fit against the gap value or the entropy values S_1 and S_0 . These are shown in Figures 6.22a-c. Figure 6.22a presents the relationship of the goodness of fit against the gap. It shows that the simultaneous methods SIM-S₁ and SIM-S₀ and the improved sequential methods ME3-S₁ and ME3-S₀ improve the goodness of fit during the the first five subproblems in which major reductions in gap are made. After the fifth subproblem, their goodness of fit gets worse, while there is no further reduction in gap. By contrast, the sequential method ME2 fails to improve the goodness of fit progressively and rather it is subject to fluctuations. Figures 6.22b and 6.22c show the relationship of the goodness of fit against the entropy values S_1 or S_0 . They show that the simultaneous methods SIM-S₁ and SIM-S₀ and the improved sequential ones ME3-S₁ and ME3-S₀ improve the goodness of fit during the first five subproblems progressively and thereafter their goodness of fit gets worse while entropy is further decreasing. The large decrease in entropy values without reducing gap during the later subproblems is due to the role of the increased penalty parameter value given to the gap penalty function. On the other hand, ME2 fails to improve the goodness of fit progressively.

(a) Goodness of fit vs Gap



(b) Goodness of fit vs Entropy(S1)

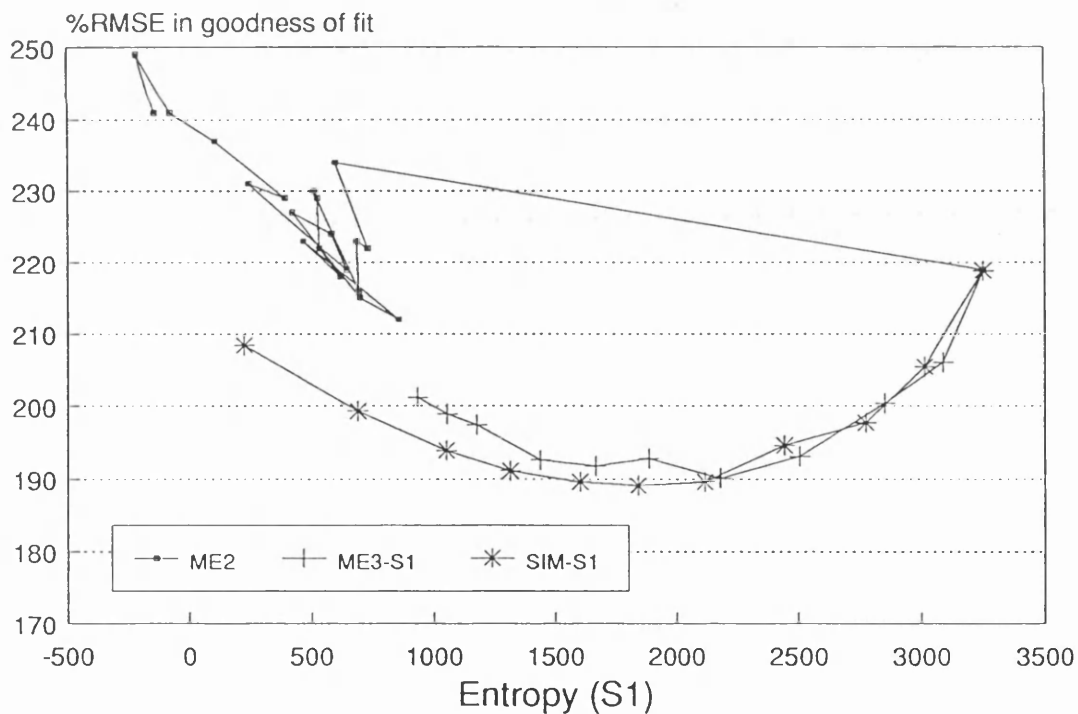


Figure 6.22 Comparisons of goodness of fit vs gap, goodness of fit vs entropy(S1), and goodness of fit vs entropy(S0)

(c) Goodness of fit vs Entropy(S₀)

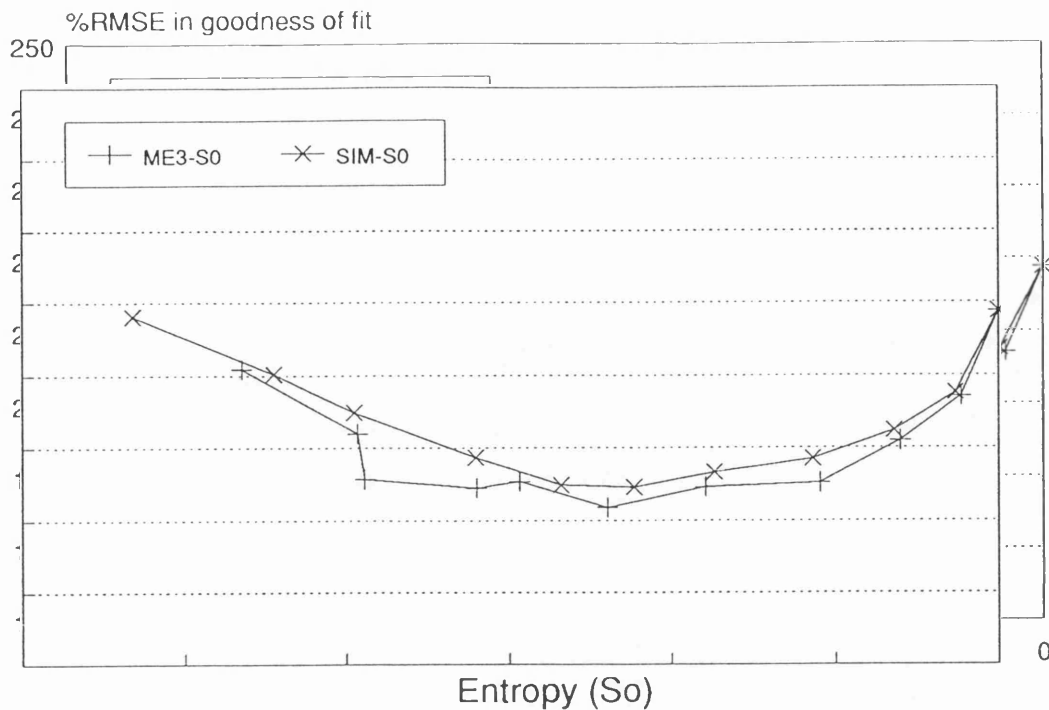


Figure 6.22 (cont.) Comparisons of goodness of fit vs gap, goodness of fit vs entropy(S₁), and goodness of fit vs entropy(S₀)

Finally, the computing times incurred to each of the alternative estimation methods are presented in Table 6.18. As shown in Table 6.18, the simultaneous methods SIM-S₁ and SIM-S₀ are not practical at all in networks such as the Ripon network used in this test, as they require about 1000 times as much as cpu time of the improved sequential methods. On the other hand, the improved sequential methods ME3-S₁ and ME3-S₀ are practical in terms of cpu times, even if they require more CPU times than does the sequential ME2 method.

Table 6.18 Computing times incurred to the alternative solution methods

Solution Methods	ME2	ME3-S ₁	SIM-S ₁	ME3-S ₀	SIM-S ₀
CPU times (seconds)	18*	30*	4500**	30*	4500**

* : On AMDAHL 5890

** : On CRAY X-MP (1 second in CRAY \cong 8 seconds in AMDAHL)

6.4.2.3 Conclusions

The following conclusions can be drawn from these results:

(1) The simultaneous methods SIM-S₀ and SIM-S₁ and the improved sequential methods ME3-S₀ and ME3-S₁ perform similarly in terms of entropy, gap, and goodness of fit. This might have important practical implications for the matrix estimation process, because the computational demands of the improved sequential methods ME3 are also practical.

(2) The simultaneous method SIM-S₁ and the improved sequential method ME3-S₁ perform better than the sequential method ME2 in terms of gap, entropy and goodness of fit.

(3) A clear trade-off curve can be identified between gap and entropy values of the trip matrices estimated for each of the proposed estimation methods SIM-S₀, SIM-S₁, ME3-S₀ and ME3-S₁. This could be a useful and practical tool for transport planners, because it allows the selection of the estimated trip matrices to be controlled depending on the relative accuracy of the prior trip matrices and observed link flows that are input. For example, convergence in gap can be a useful stopping criteria for selecting a trip matrix. In the results of this test, a most appropriate trip matrix is obtained at the fifth or sixth subproblems at which convergence in gap is somewhat achieved.

(4) The full entropy objective function $S_0(\underline{T}, t)$ tends to estimate trip matrices whose totals are larger than those by the simplified one $S_1(\underline{T}, t)$ during the sequence. However, their difference in respect of other performance indicators is shown to be not substantial.

6.4.3 Tests with real link flows

6.4.3.1 Introduction

In the previous section, error-free link flows obtained by assigning a known trip matrix to the network were used to estimate a trip matrix. In this section, the real link flows observed as traffic

counts collected in the town of Ripon were used to estimate a trip matrix. The main objective of this test is to investigate how well the estimation methods can deal with inconsistent link flows in a real network. The proposed estimation methods SIM and ME3 are expected to be more flexible in accommodating inconsistent link flows than is the sequential method ME2. However, as no observations were available of a trip matrix which corresponds to the traffic counts used in this test, it was not possible to carry out any calculation of the goodness of fit between the estimated trip matrices and the observed one. A uniform prior trip matrix with cell values t_{ij} each equal to 5 was used as a prior trip matrix.

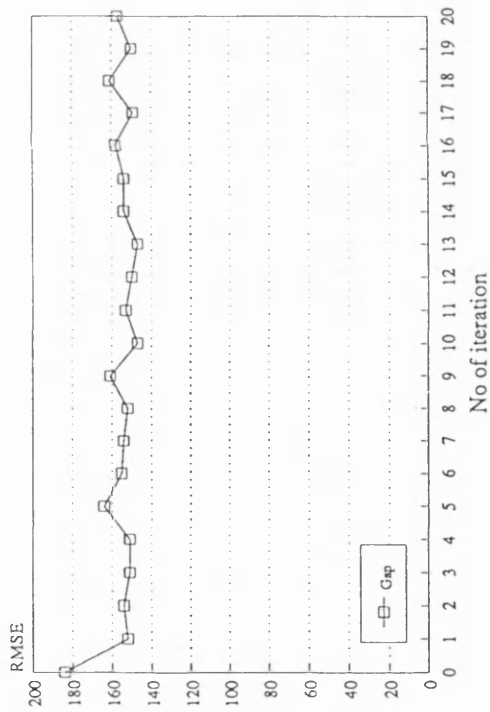
6.4.3.2 Results

The results of the test using the sequential method of ME2 - totals of estimated trip matrices, entropy values and gap values - are summarized in Table 6.19. The evolution of these indicators is depicted graphically in Figures 6.23a-c.

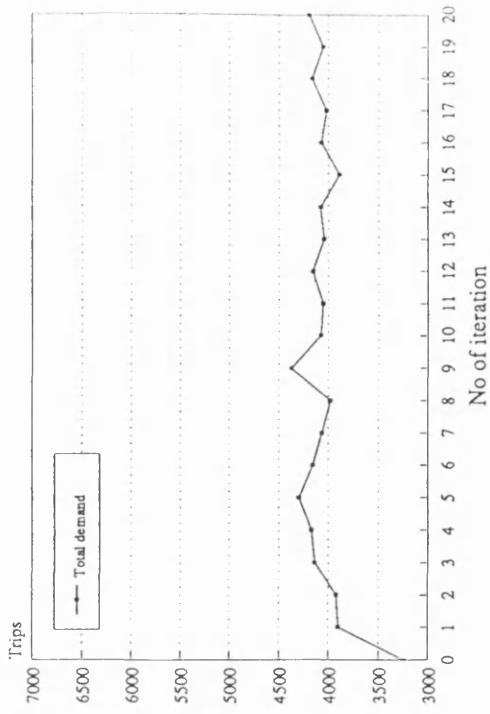
Table 6.19 Results of the test using real link flows by ME2

Iteration	T..	Entropy		Gap	
		S0	S1	RMSE	%RMSE
Observed	3949	-4111	-67	168	70
Prior	3250	0	3250	184	77
1	3900	-3995	-799	152	63
2	3920	-4470	-1204	154	64
3	4132	-5462	-2290	151	63
4	4166	-4560	-1291	151	63
5	4295	-5555	-2215	164	68
6	4153	-5888	-2583	155	65
7	4063	-5075	-1774	154	64
8	3977	-4412	-1149	152	63
9	4371	-5173	-2041	161	67
10	4073	-4732	-1458	147	61
11	4052	-4251	-916	153	64
12	4156	-4696	-1481	150	63
13	4042	-5106	-1857	147	61
14	4079	-4407	-1231	154	64
15	3888	-4537	-1282	154	64
16	4070	-4271	-1053	158	66
17	4018	-4670	-1481	149	62
18	4163	-4756	-1567	161	67
19	4050	-4812	-1565	150	63
20	4190	-4866	-1628	157	66

(a) Evolution of gap



(b) Evolution of total demand



(c) Trade-off curve

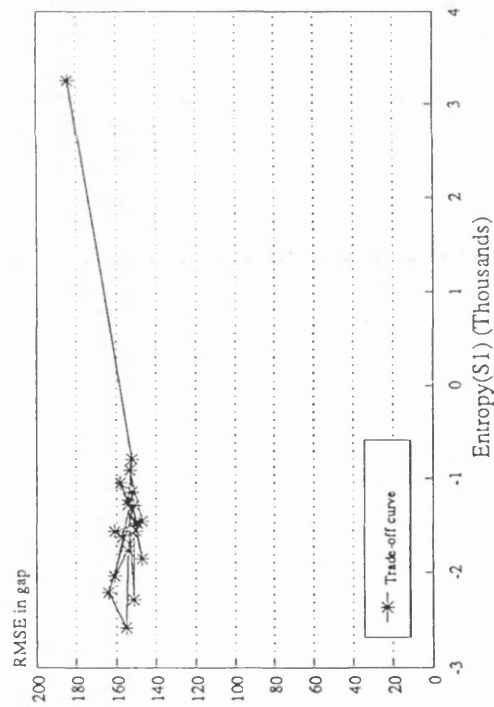


Figure 6.23 Results of the test using real link flows by ME2

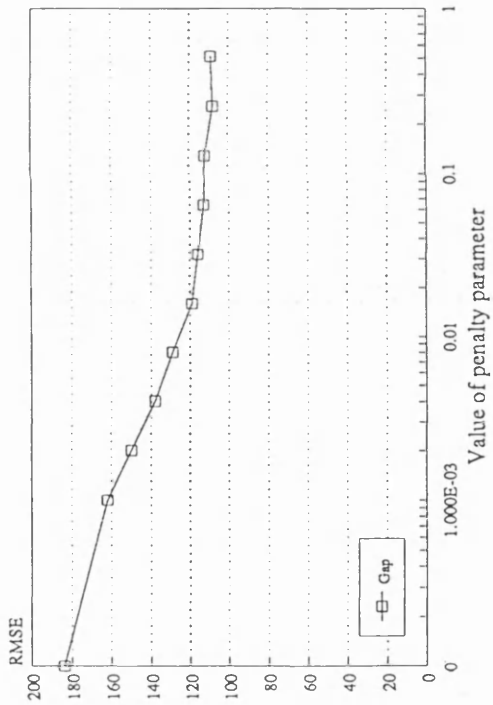
Figure 6.23a shows that ME2 reduces gap rapidly during the first iteration and makes considerable fluctuations thereafter. A similar performance to reduction in gap was obtained for the totals of the estimated trip matrices, as is shown in Figure 6.23b. Figure 6.23c shows the trade-off relationship between entropy values and gap values of estimated trip matrices. It shows that it is difficult to identify any clear trade-off relationship after the first iteration.

The results of the test using the simultaneous method SIM-S₁ are presented in Table 6.20 and depicted graphically in Figures 6.24a-c. Figure 6.24a shows that as the magnitude of the penalty parameter increases, SIM-S₁ reduces the gap value gradually, approaching the RMSE value in gap of 109 (46%). Also, it shows that a relatively small amount of reduction in gap is made by comparison with the large decrease in entropy. As shown in Figures 6.24b, SIM-S₁ fails to converge to any particular value of total demand and this is due to the rapid increase in the penalty parameter value.

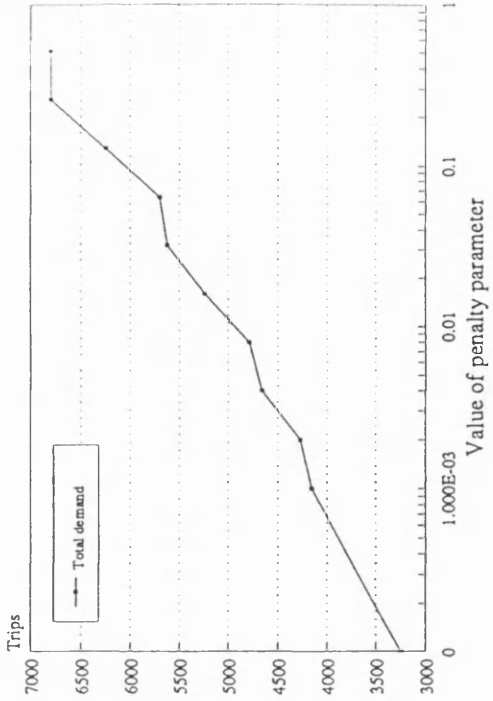
Table 6.20 Results of the test using real link flows by SIM-S₁

Penalty Parameter	T..	Entropy		Gap	
		S0	S1	RMSE	%RMSE
0.000	3250	0	3250	184	77
0.001	4153	-396	2739	162	68
0.002	4271	-851	2253	150	63
0.004	4660	-1666	1315	138	58
0.008	4787	-2734	198	129	54
0.016	5239	-3974	-1236	119	50
0.032	5622	-5097	-2556	116	48
0.064	5703	-5659	-3163	113	47
0.128	6245	-6894	-4728	112	47
0.256	6802	-7964	-6186	108	45
0.512	6804	-8360	-6583	109	45

(a) Evolution of gap



(b) Evolution of total demand



(c) Trade-off curve

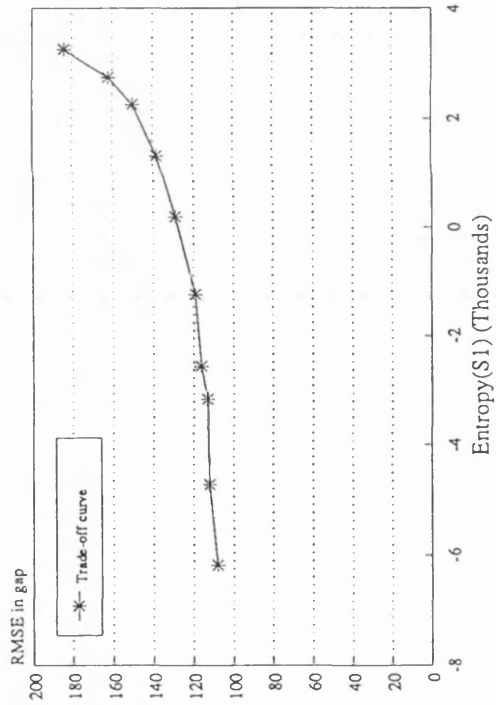


Figure 6.24 Results of the test using real link flows by SIM-S1

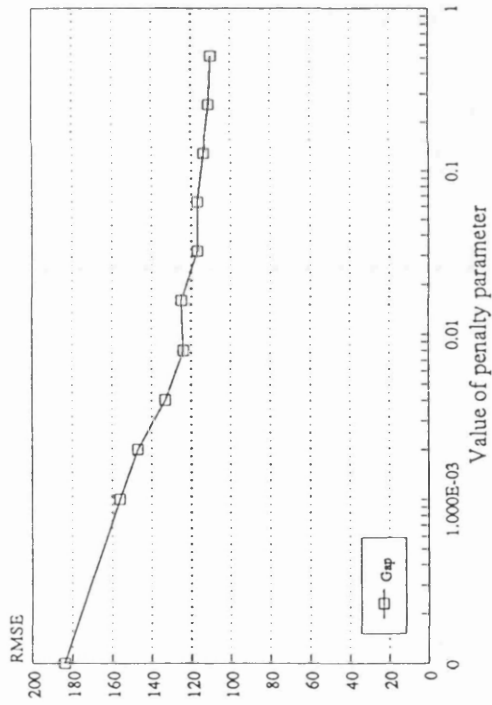
Figures 6.24c shows that a clear trade-off curve between gap values and entropy values exist. It is interesting to note that after the first six subproblems the entropy value is further reduced without any substantial reduction in the gap.

A similar result to the one by SIM-S₁ was obtained by the simultaneous method SIM-S₀. The results are presented in Table 6.21 and depicted graphically in Figures 6.25a-c. As shown in Figures 6.24b and 6.25b, the main difference in the performance between SIM-S₁ and SIM-S₀ lies in the totals of the estimated trip matrices, as observed in the results from other tests.

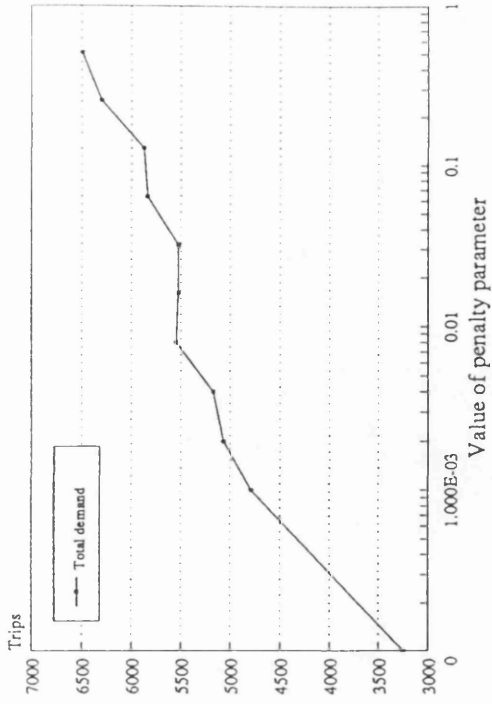
Table 6.21 Results of the test using real link flows by SIM-S₀

Penalty Parameter	T..	Entropy		Gap	
		S0	S1	RMSE	%RMSE
0.000	3250	0	3250	184	77
0.001	4788	-401	2532	156	65
0.002	5069	-976	1840	147	61
0.004	5173	-1959	810	133	56
0.008	5549	-3123	-543	124	52
0.016	5527	-3789	-1196	125	52
0.032	5525	-4927	-2334	117	50
0.064	5834	-5524	-3103	117	49
0.128	5872	-6491	-4093	114	48
0.256	6298	-7116	-4983	111	46
0.512	6491	-7670	-5669	110	46

(a) Evolution of gap



(b) Evolution of total demand



(c) Trade-off curve

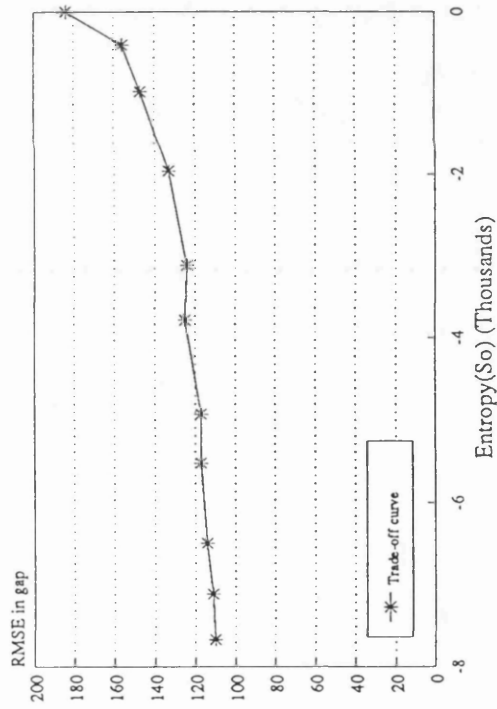


Figure 6.25 Results of the test using real link flows by SIM-So

Similar results to the ones of the simultaneous methods were produced by the improved sequential methods ME3-S₁ and ME3-S₀. The results were presented in Tables 6.22 and 6.23 and depicted graphically in Figures 6.26a-c and 6.27a-c. Figures 6.26a and 6.27a show that as the magnitude of the penalty parameter increases, ME3-S₁ and ME3-S₀ reduce the gap value. They show that ME3-S₀ performs better in the reduction in gap than does ME3-S₁. As shown in Figures 6.26b and 6.27b, an examination of the totals of the estimated trip matrices shows that neither ME3-S₁ or ME3-S₀ converge to any particular value, as was observed in the results of the simultaneous solution methods. Figures 6.26c and 6.27c show trade-off curves between the gap values and entropy values for the estimated trip matrices. They show that there exist some small fluctuations in the trade-off curves.

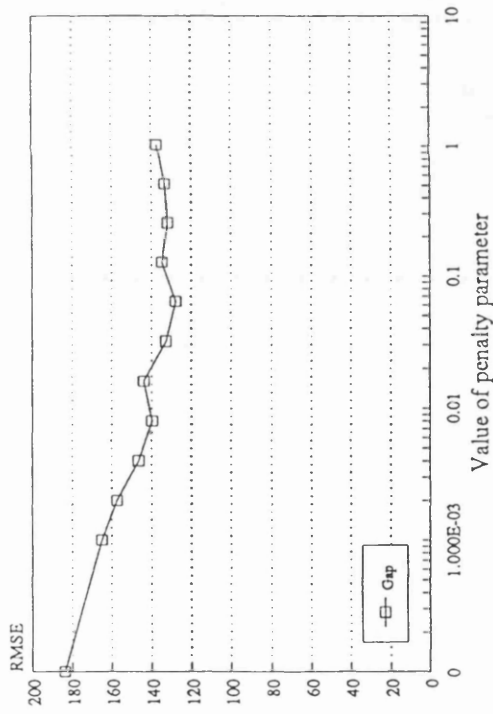
Table 6.22 Results of the test using real link flows by ME3-S₁

Penalty Parameter	T..	Entropy		Gap	
		S0	S1	RMSE	%RMSE
0.000	3250	0	3250	184	77
0.001	3723	-231	2986	165	69
0.002	3690	-792	2429	157	66
0.004	3696	-1633	1588	147	61
0.008	3999	-2723	447	140	58
0.016	4236	-3636	-522	144	60
0.032	4405	-5342	-2276	133	55
0.064	4699	-6429	-3462	128	53
0.128	4878	-7123	-4226	135	56
0.256	4829	-7312	-4395	132	55
0.512	4829	-7206	-4289	133	56

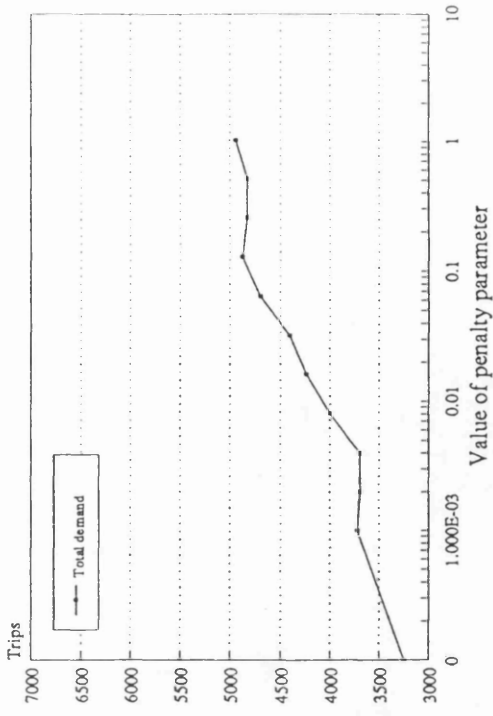
Table 6.23 Results of the test using real link flows by ME3-S₀

Penalty Parameter	T..	Entropy		Gap	
		S0	S1	RMSE	%RMSE
0.000	3250	0	3250	184	77
0.001	3946	-235	2945	163	68
0.002	3936	-896	2287	158	66
0.004	3876	-1570	1623	149	62
0.008	4082	-2709	442	143	60
0.016	4353	-3682	-601	134	56
0.032	4467	-4941	-1895	140	58
0.064	4475	-5482	-2438	139	58
0.128	4856	-7028	-4122	129	54
0.256	5159	-7959	-5184	119	50
0.512	5548	-8401	-5819	121	51

(a) Evolution of gap



(b) Evolution of total demand



(c) Trade-off curve

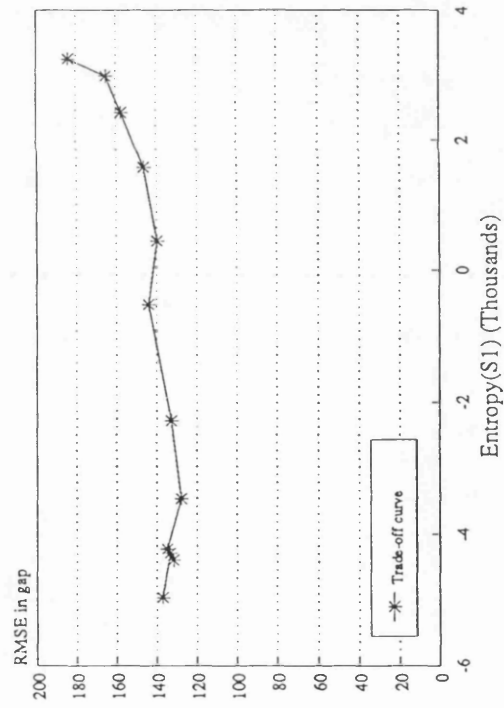
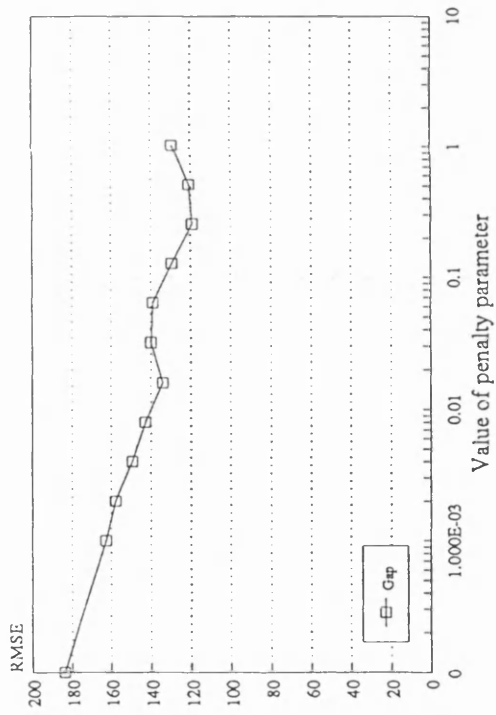
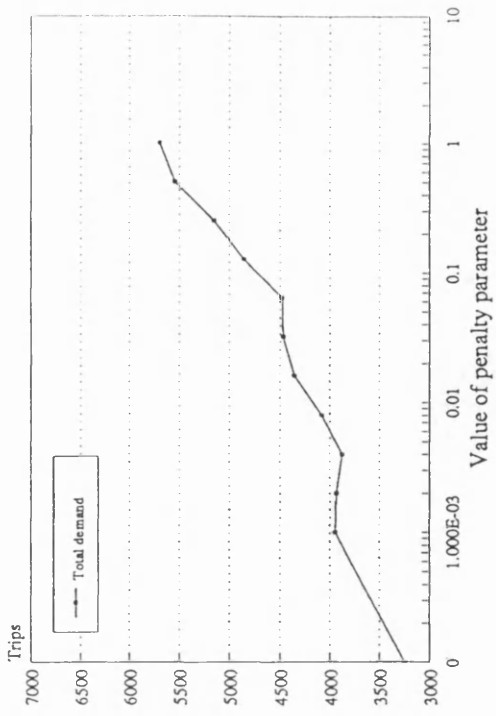


Figure 6.26 Results of the test using real link flows by ME3-S1

(a) Evolution of gap



(b) Evolution of total demand



(c) Trade-off curve

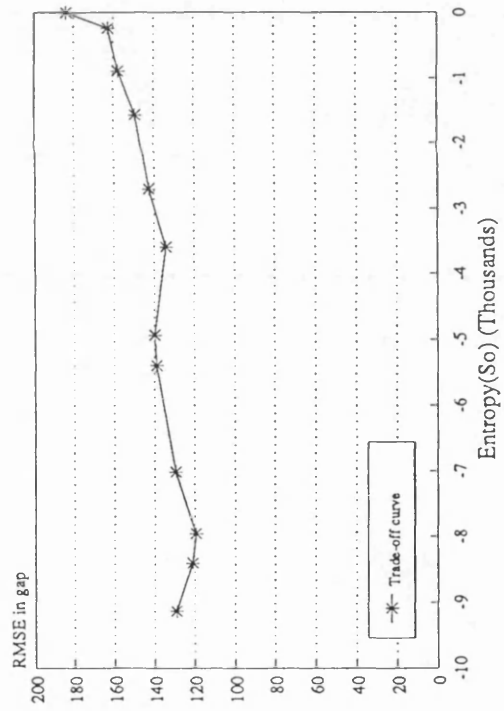
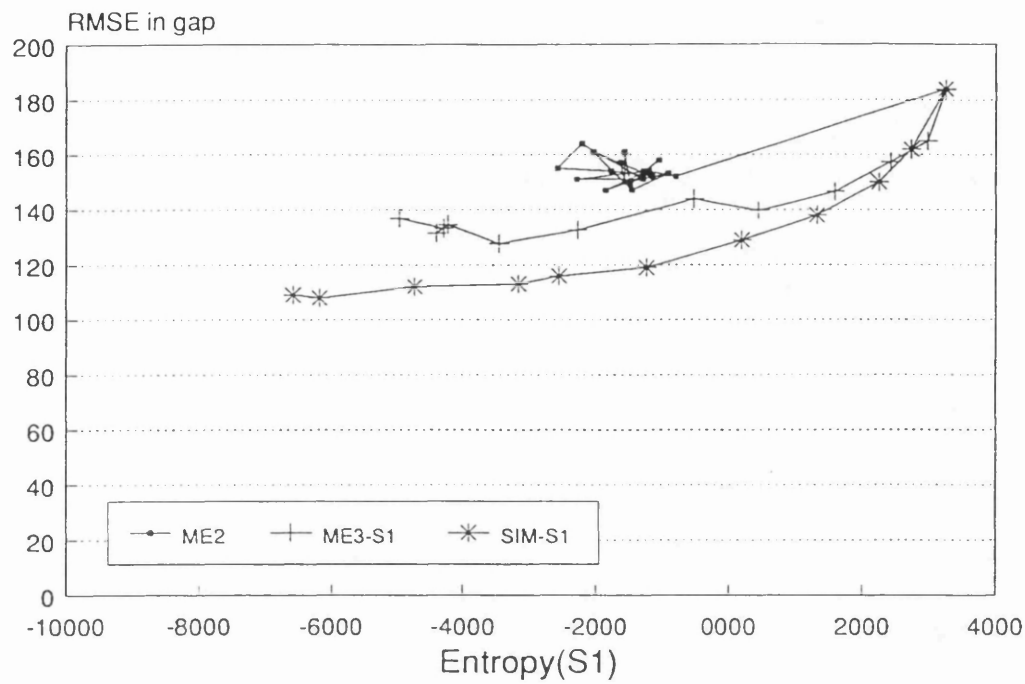


Figure 6.27 Results of the test using real link flows by ME3-So

For comparison purposes, the trade-off curves of three alternative estimation methods ME2, ME3-S₁ and SIM-S₁ are shown together in Figure 6.28a and those of ME3-S₀ and SIM-S₀ are shown together in Figure 6.28b.

(a) By the methods using the entropy function $S_1(T,t)$



(b) By the methods using the entropy function $S_0(T,t)$

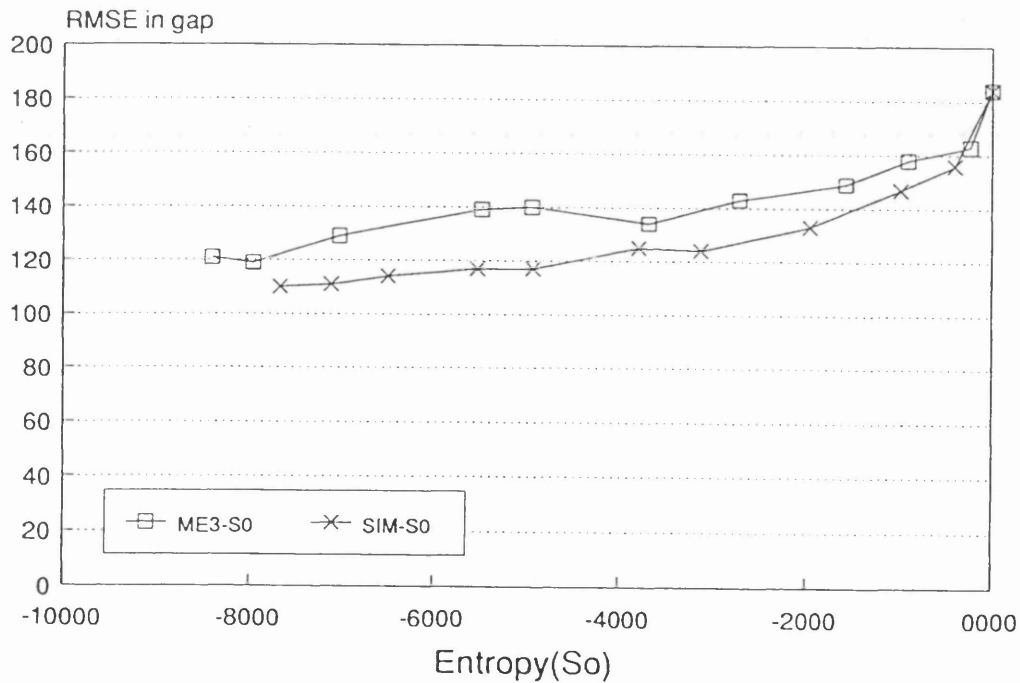


Figure 6.28 Comparison of the trade-off curves from real link flows by various solution methods

Figure 6.28a shows that SIM-S₁ performs better than do ME3-S₁ and ME2. The performance of ME3-S₁ is not bad, although ME3-S₁ performs worse than SIM-S₁. ME2 performs worst compared with the other two methods that use S₁(T,t) as an objective function. This result contrasts with the result using the error-free modelled link flows in which the simultaneous method and the improved sequential method perform very closely. This suggests that the simultaneous methods deal with inconsistent link flows better than do the improved sequential methods or the sequential ME2 method. A similar comment to the result using S₁(T,t) can be made on the result using the full entropy function S₀(T,t). As shown in Figure 6.28b, SIM-S₀ performs somewhat better than ME3-S₀.

6.4.3.3 Conclusions

The following conclusions can be drawn from these results.

(1) The simultaneous methods SIM-S₀ and SIM-S₁ perform best amongst all the estimation methods. It is interesting to note that the simultaneous methods SIM-S₁ and SIM-S₀ perform much better than the improved sequential methods ME3-S₁ and ME3-S₀, especially when matrix is estimated from inconsistent real traffic counts. The performance of the improved sequential methods ME3-S₁ and ME3-S₀ is not bad, although they perform worse than the simultaneous ones. However, the sequential ME2 method fails to reduce the gap after the first one or two iterations when inconsistent link flows are used.

(2) When inconsistent link flows are used to estimate a trip matrix, the trade-off curves between the gap values and the entropy ones produced by the simultaneous methods or the improved sequential ones could be even more useful in the matrix estimation exercise than when a trip matrix is estimated from error-free link flows, as these curves allow us to select a trip matrix depending on the relative accuracy of the prior trip matrix and traffic counts that are input. If we apply the same stopping criterion - convergence in gap - as mentioned in Section 6.4.2.3 to the results of this test, for example an appropriate trip matrix estimated

by the simultaneous method SIM-S₁ is obtained at the sixth subproblem as convergence in gap is achieved.

(3) Use of the full entropy objective function $S_0(\underline{T}, \underline{t})$ tends to give rise to estimates of trip matrices whose totals are bigger than those estimates by comparable methods which use the simplified one $S_1(\underline{T}, \underline{t})$. However, the difference in other performance indicators is shown to be not substantial.

CHAPTER 7.

CONCLUSIONS AND SUGGESTIONS FOR FURTHER STUDY

This chapter is organized as follows. Section 7.1 draws the conclusions from the main findings of this research. Section 7.2 suggests some areas for further research.

7.1 Conclusions

The estimation of trip matrices is an important part in the analysis of traffic management and transport planning tasks. The trip matrices are often used to help design and evaluate new transport plans.

Conventional methods for estimating trip matrices appear to be inaccurate, disruptive and expensive. It was recognized that the use of traffic counts for estimating trip matrices can avoid at least some of the difficulties identified in conventional methods. In particular, traffic counts are relatively convenient and inexpensive to collect, and moreover they can be used to update old trip matrices.

According to a review of the literature, the most common idea of estimating trip matrices from traffic counts is to find a trip matrix which, when assigned to the network closely reproduces some observed traffic counts. Three fundamental issues associated with the estimation problem were identified. These were: underspecification of the trip matrix, inconsistency between traffic counts, and congestion effects. Each of these requires a special treatment in the matrix estimation method. In particular, the use of an appropriate traffic assignment method is important in the matrix estimation process, especially when congestion in networks plays an important role in route choice.

The main objective of this study was to develop new methods for estimating trip matrices from traffic counts when congestion effects in networks are considered.

This study proposed a new formulation and solution method which solves the two subproblems of equilibrium assignment and entropy maximization simultaneously. The new formulation uses equilibrium link flows in the constraints rather than route choice proportions. One of a range of objective functions can be adopted and two different measures of entropy have been used.

As a solution method, the new formulation was first transformed into a sequence of unconstrained optimization problems by using a penalty function method. A heuristic method was developed to approximate the equilibrium link flows by fitting linear functions to represent the variation in link flows with respect to changes in trip matrix elements. This was found to be useful because it helps to reduce the computational burden and also to overcome the non-convexity of the equilibrium constraints. However, it was recognized that the goodness of fitting the approximate linear relationship over sets of equilibrium link flows depends on the convergence to the equilibrium link flows from the equilibrium assignment.

In spite of the use of an extrapolation method and a perturbation method, it was found that the simultaneous method is impractical for use in large networks because of its considerable computational requirements. An improved sequential method which uses a penalty function approach was therefore proposed. This method approximates an optimum solution by approaching the feasible region progressively. Fixed route choice proportions are used within each matrix estimation subproblem and are updated by the equilibrium assignment subproblem.

The proposed estimation methods have been tested and their performances have been compared with that of the sequential ME2 one. Three tests using both artificial and real networks have been carried out and the main findings from each of these three tests have been reported in detail. Here, we draw some general conclusions from those findings.

(1) The simultaneous method SIM works successfully and performs better, especially in congested networks, than either the sequential method ME2 or the improved sequential method ME3 with respect to gap, entropy and goodness of fit. However, the computational complexity of the simultaneous method appears to be its main drawback and its application to real estimation problems is not practical with current computer technology. However, computing capability is improving rapidly. This advance in the computing technology will permit the simultaneous method to be more practical in the future.

(2) The improved sequential method ME3 performs closely to the simultaneous one SIM in terms of gap, entropy and goodness of fit. It performs better than the sequential method ME2. The improved sequential method ME3 is practical with respect to the computational demand, even if they require more cpu times than does ME2. The use of the improved sequential method ME3 is therefore recommended.

(3) Use of the full entropy function $S_0(\underline{T}, \underline{t})$ appears to give rise to estimated trip matrices which have totals that are relatively bigger than those estimated using the simplified one $S_1(\underline{T}, \underline{t})$. Despite the theoretical preference for $S_0(\underline{T}, \underline{t})$, the difference in other performance indicators such as gap and goodness of fit in congested networks is not substantial.

(4) A clear trade-off curve can be identified between entropy and gap during the solution process of the simultaneous method SIM and the improved sequential method ME3. This could be a useful and practical tool for transport planners because it allows the selection of estimated trip matrices to be controlled depending on the relative accuracy of the prior trip matrices and traffic counts that are used as input.

(5) The new formulation and solution method - the improved sequential solution method ME3 - can be equally applied to the matrix estimation in uncongested networks without any modifications. The new method is more flexible to deal with inconsistent traffic counts and it does not require any prior corrections to inconsistent traffic counts. Moreover, the trade-off curves between entropy and gap values could be useful for

the selection of a trip matrix.

(6) An interesting feature of the new formulation and solution methods is that they are modular and so they can accommodate without any modifications objective functions or penalty functions other than the ones used in this study. Some suggestions for other objective functions or penalty functions will be given in Section 7.2.

(7) Despite its theoretical weakness, the sequential method ME2 has been being used in the field because of its practical advantage in terms of computing time. The evidence discovered in this study indicates that in several cases the sequential method ME2 either fails to converge to an optimal solution or cannot improve the goodness of fit of the estimated trip matrices. Its performance is subject to considerable fluctuations. However, where it is used, examination of its performance indicators such as entropy and gap could be useful for both the decision of the number of iterations required and the selection of an estimated trip matrix.

7.2 Suggestions for further study

During this research, a number of areas for further research have been identified. The most important ones among these are outlined below.

(1) **Further tests with real data sets:** One of the main difficulties of using the Ripon data base was that a real peak-hour trip matrix observed independently from traffic counts is not available for the present study. Consequently, it was not possible for this study to perform an ideal test in which the trip matrices estimated from traffic counts are compared to the real observed trip matrix. Further tests using the real observed trip matrix and traffic counts are required to see how well the proposed methods perform in practice.

(2) **Extension for other objective functions:** Although the entropy functions are useful as objective functions, it is also possible to accommodate other objective functions such as a quadratic function used

by McNeil and Hendrickson (1984a), Cascetta (1984) and LeBlanc and Farhangian (1982) and a likelihood function used by Spiess (1987).

(3) **Extension for other penalty functions:** The gap penalty function used in this study can be replaced with a likelihood function. A likelihood function can be derived by assuming that observed flows follow a Poisson distribution. This gives a log-likelihood function:

$$L(\underline{V}^*(\underline{T})|\underline{V}) = \sum_{a \in I} (\bar{V}_a - V_a^*) + V_a \log_e (V_a^*/\bar{V}_a) \quad (7.1)$$

The log-likelihood function $L(\underline{V}^*(\underline{T})|\underline{V})$ is less than or equal to 0 and has value 0 iff $\bar{V}_a = V_a^*$ for all a and is continuous. Furthermore, it satisfies all the properties required of a penalty function. One main feature of maximizing this likelihood function is that it does not restrict the search to strictly feasible solutions.

(4) **Application for the network design problem:** The problem of designing elements of a road network while traffic flows vary corresponding to any design changes is known as the equilibrium network design problem. The network design problem can be viewed as a bi-level decision making problem (Heydecker, 1986): at the upper level, planners seek to optimize the operational performance of the network, whilst at the lower level, individual drivers make choices with regard to route, mode, origin and destination which they perceive to be best for themselves. The network design problem can be expressed as a similar mathematical formulation to the matrix estimation problem in which two subproblems of equilibrium assignment and matrix estimation are solved simultaneously. A method of constraint approximation of equilibrium link flows developed in this study can also be applied to transform the network design problem into a sequence of sub-problems with linear constraints (Heydecker and Khoo, 1990).

REFERENCES

- AREZKI, Y. and VAN VLIET, D. (1985) The use of quantal loading in equilibrium traffic assignment. Transportation Research, 19B(6), 521-525.
- AREZKI, Y. and VAN VLIET, D. (1990) A full analytical implementation of the PARTAN/Frank-Wolfe algorithm for equilibrium assignment. Transportation Science, 24(1), 58-62.
- ARMSTRONG, B.D. (1977) The need for route guidance. Supplementary Report. 330, TRRL.
- ATKINS, S.T. (1987) The crisis for transportation planning modelling. Transport Review, 7(4), 307-325.
- BATTY, M. and MARCH, L. (1976) The method of residues in urban modelling. Environment and Planning A, 8, 189-214.
- BECKMANN, M., McGUIRE, C.B. and WINSTEN, C. (1956) Studies in the economics of transportation. Cowles Commission Monograph, Yale University Press, New Haven.
- BELL, M.G.H. (1983) The estimation of an origin-destination matrix from traffic counts, Transportation Science, 17(2), 198-217.
- BELL, M.G.H. (1984) Log-linear models for the estimation of origin-destination matrices from traffic counts: An approximation. Proceedings of the Ninth International Symposium on Transportation and Traffic Theory, Delft, 451-469.
- BELL, M.G.H. (1985) Variances and covariances for origin-destination flows when estimated by log-linear models. Transportation Research, 19B(6), 497-507.

- BELL, M.G.H (1991) The estimation of origin-destination matrices by constrained generalized least squares. Transportation Research, 25B(1), 13-22.
- BENSHOOF, J.A. (1970) Characteristics of drivers' route selection behaviour. Traffic Engineering and Control, 11, 604-606.
- BREGMAN, G.C. (1967) The relaxation method of finding the common points of convex sets and its application to the solution of problems in convex programming. USSR Computational Mathematical Physics, 7, 200-217.
- BRENNINGER-GÖTHE, M., JÖRNSTEN, K.O. and LUNDGREN, J.T. (1988) Estimation of origin-destination matrices from traffic counts using multi-objective programming formulations. Publication 88-14, Department of Mathematics, University of Linköping.
- BRILLOUIN, L. (1956) Science and Information Theory. Academic Press, New York.
- BROOKS, I.J. and HARRIS, M.R. (1972) The application of multi-flow assignment to major traffic studies in a rural area. Traffic Engineering and Control, 14, 330-332.
- BURRELL, J.E. (1968) Multiple route assignment and its applications to capacity restraint. Proceedings of the 4th International Symposium on the Theory of Traffic Flow, Karlsruhe, 1968.
- CAREY, M., HENDRICKSON, C. and SIDDHARTHAN, K. (1981) A method for direct estimation of origin-destination matrices. Transportation Science, 15(1), 32-49.
- CAREY, M. and REVELLI, R. (1986) Constrained estimation of direct-demand functions and trip matrices. Transportation Science, 20(3), 143-152.

- CASCETTA, E. (1984) Estimation of trip matrices from traffic counts and survey data: A generalized least-squares estimator. Transportation Research, 18B(4/5), 289-299.
- CASCETTA, E. and NGUYEN, S. (1988) A unified framework for estimating or updating origin/destination matrices from traffic counts. Transportation Research, 22B(6), 437-455.
- DAFERMOS, S.C. and SPARROW, F.T. (1969) The traffic assignment problem for a general network. National Bureau of Standards J. Res., 73B, 91-118.
- DAGANZO, C.F and SHEFFIC, Y. (1977) On stochastic models of traffic assignment. Transportation Science, 11, 253-274.
- DEPARTMENT OF TRANSPORT (1981). Traffic Appraisal Manual.
- DIAL, R.B. (1971) A probabilistic multipath traffic assignment model which obviates path enumeration. Transportation Research, 5(2), 83-111.
- DOMENCICH, T.A., KRAFT, G. and VALLETTE, J. (1968) Estimation of urban travel passenger travel behaviour: An economic demand model. Highway Research Record, 238, 64-78.
- DOMENCICH, T.A. and McFADDEN, D. (1975) Urban Travel Demand: A Behavioral Analysis. North-Holland Publishing Company, Amsterdam.
- EASH, R.W., JANSON, B.N. and BOYCE, D.E. (1979) Equilibrium trip assignment: advantages and implications for practice. Transportation Research Board, 728, 1-8.
- ERLANDER, S. (1977) Accessibility, entropy and the distribution and assignment of traffic. Transportation Research, 12(5), 149-153.

- ERLANDER, S., NGUYEN, S. and STEWART, N.F. (1979) On the calibration of the combined distribution-assignment model. Transportation Research, 13B, 259-267.
- EVANS, S.P. (1976) Derivation and analysis of some models for combining trip distribution and assignment. Transportation Research, 10, 37-57.
- EVANS, S.P. and KIRBY, H.R. (1974) A three-dimensional Furness procedure for calibrating gravity models. Transportation Research, 8(2), 105-122.
- FIACCO, A.V. and McCORMICK, G.P. (1968) Nonlinear programming: sequential unconstrained minimization techniques. New York: John Wiley and Sons.
- FISK, C.S. (1984a) Game theory and transportation systems modelling. Transportation Research, 18B(4/5), 301-313.
- FISK, C.S. (1984b) Optimal signal controls on congested networks. Proc. of the 9th International Symposium on Transportation and Traffic Theory, Delft, 199-216.
- FISK, C.S. (1988) On combining maximum entropy trip matrix estimation with user optimal assignment. Transportation Research, 22B(1), 69-79.
- FISK, C.S. (1989) Trip matrix estimation from link traffic counts: the congested network case. Transportation Research, 23B(5), 331-336.
- FISK, C.S. and BOYCE, D.E. (1983) A note on trip matrix estimation from link traffic count data. Transportation Research, 17B(3), 245-250.
- FLORIAN, M. (1977) An improved linear approximation algorithm for the network equilibrium problem. Publication no. 57, Centre de Recherche sur les Transports, Université de Montréal.

- FLORIAN, M. and FOX, B. (1976) On the probabilistic origin of Dial's multipath traffic assignment model. Transportation Research, 10(5), 339-341.
- FLORIAN, M. and NGUYEN, S. (1976) An application and validation of equilibrium assignment methods. Transportation Science, 10(4), 374-390.
- FLORIAN, M., NGUYEN, S. and FERLAND, J. (1975) On the combined distribution assignment of traffic. Transportation Science, 9, 43-53.
- FRANK, M. and WOLFE, P. (1956) An algorithm for quadratic programming. Naval Research Logistics, Quarterly, 3, 95-110.
- FUKUSHIMA, M. (1984) A modified Frank-Wolfe algorithm for solving traffic assignment problem. Transportation Research, 18B(2), 169-177.
- FURNESS, K.P (1965) Time function iteration. Traffic Engineering and Control, 7(7), 458-460.
- GUÉLAT, J. and MARCOTTE, P. (1986) Some comments on Wolfe's 'away step'. Mathematical Programming, 35, 110-119.
- GUR, Y., TURNQUIST, M.A., SCHNEIDER, M., LeBLANC, L., and KURTH, D. (1979) Estimation of an origin-destination trip table based on observed link volumes and turning movements. Final Report, Number DOT-FH-11-9292, U.S. federal Highway Administration, Washington, D.C.
- HALL, M.D., VAN VLIET, D. and WILLUMSEN, L.G. (1980) SATURN - a simulation-assignment model for the evaluation of traffic management schemes. Traffic Engineering and Control, 21(4), 168-176.

- HEYDECKER, B.G. (1986) On the definition of traffic equilibrium. Transportation Research, 20B(6), 435-440.
- HEYDECKER, B.G. (1986) The equilibrium network design problem: a critical review. Transport Studies Group, University College London. (unpublished)
- HEYDECKER, B.G. and KHOO, T.K. (1990) The equilibrium network design problem. Proc. of the 22nd UTSG Annual Conference, Hatfield Polytechnic, London, 1990. (unpublished)
- HÖGBERG, P. (1976) Estimation of parameters in models for traffic prediction: A non-linear-regression approach. Transportation Research, 10(4), 263-265.
- HOLM, J., Jensen, T., Nielsen, S., Christensen, A., Johnsen, B. and Ronby, G. (1976) Calibrating traffic models on traffic census only. Traffic Engineering and Control, 17(4), 137-140.
- INSTITUTION OF HIGHWAYS AND TRANSPORTATION AND DTp (1987). Roads and Traffic in Urban Areas. London: HMSO.
- ISAACSON, E. and KELLER, H.B. (1966) Analysis of Numerical Methods. John Wiley & Sons, Inc., New York.
- JAYNES, E.T. (1957) Information theory and statistical mechanics. Physical Review, 106, 620-630; and 108, 171-190.
- JÖRNSTEN, K. and NGUYEN, S. (1979) On the estimation of a trip matrix from network data. Publication no. 153, Centre de Recherche sur les Transports, Université de Montréal.
- KIRBY, H.R. (1979) Partial matrix technique. Traffic Engineering and Control, 20(8/9), 422-428.

- KRUIITHOF, J. (1937) Calculation of telephone traffic. De Ingenieur, 52(8), E15-E25. English translation by UK Post Office Research Department Library (No. 2663), Dollis Hill, London.
- LAM, H.K. and LO, H.P. (1990) Accuracy of O-D estimates from traffic counts. Traffic Engineering and Control, 31(6), 358-367.
- LAMOND, B. and STEWART, N.F. (1981) Bregman's balancing method. Transportation Research, 15B(4), 239-248.
- LeBLANC, L. and FARHANGIAN, K.(1982) Selection of a trip matrix which reproduces observed link flows. Transportation Research, 16B(2), 83-88.
- LeBLANC, L., HELGASON, R.V. and BOYCE, D.E. (1982) Improved efficiency of the Frank-Wolfe algorithm. Working Paper, 88-131, Vanderbilt University, Nashville, Tennessee.
- LeBLANC, L., MORLOK, E.K. and PIERSKALLA, W.P. (1975) An efficient approach to solving the road network equilibrium traffic assignment problem. Transportation Research, 9, 309-318.
- LEONARD, D.R. and TOUGH, J.B. (1979) Validation work on CONTRAM - a model for use in the design of traffic management schemes. Proc. of PTRC Summer Annual Meeting, University of Warwick, July, 1979. Seminar J, 135-153.
- LEVENTHAL, L.U., NEMHAUSER, G.L. and TROTTER, L.E. (1973) A column generation algorithm for optimal traffic assignment. Transportation Science, 7(2), 168-176.
- LOW, D.E. (1972) A new approach to transportation systems modelling. Traffic Quarterly, 26(3), 391-404.
- LUENBERGER, D.G. (1984) Linear and nonlinear programming. Reading, MA: Addison-Wesley Publishing Company.

- LUMM, S.E. (1978) Route choice by drivers. Supplementary Report 374, TRRL.
- MAHER, M.J. (1983) Inferences on trip matrices from observations on link volumes: A Bayesian statistical approach. Transportation Research, 17B(6), 435-477.
- MAHER, M.J. (1987) Bias in the estimation of O-D flows from link counts. Traffic Engineering and Control, 28(12), 624-627.
- MARCOTTE, P. (1983) Network optimization with continuous control parameters. Transportation Science, 17, 181-197.
- MASON, P.R. (1972) Multiflow multiple routing procedure. Traffic Engineering and Control, 14, 280-282.
- MATZOROS, T., VAN VLIET, D., RANDLE, J. and WESTON, B. (1987) A validation of the SATURN and ME2 models using before-and-after survey data from Manchester. Traffic Engineering and Control, 2(12), 641-643.
- McNEIL, S. and HENDRICKSON, C. (1985a) A regression formulation of the matrix estimation problem. Transportation Science, 19(3), 278-292.
- McNEIL, S. and HENDRICKSON, C. (1985b) A note on alternative matrix entry estimation techniques. Transportation Research, 19B(6), 509-519.
- MURCHLAND, J.D. (1977) The multi-proportional problem. University College London, Transport Studies Group. Note JDM-263. (unpublished)
- MURCHLAND, J.D. (1978) Applications, history and properties of bi- and multi-proportional models. Seminar at the London School of Economics, London. (unpublished)

- NEFFENDORF, H. and WOOTTON, H.J. (1974) A travel estimation model, based on screen-time interviews. Proc. of the 4th PTRC Summer Annual Meeting, July, 1974.
- NGUYEN, S. (1974a) Traffic equilibrium methods. Proc. of the International Symposium held at the Université de Montréal, Edited by M.A. Florian, November, 1974.
- NGUYEN, S. (1974b) An algorithm for the traffic assignment problem. Transportation Science, 8(3), 203-216.
- NGUYEN, S. (1976) Procedures for equilibrium traffic assignment with elastic demand. Publication no. 39, Centre de Recherche sur les Transports, Université de Montréal.
- NGUYEN, S. (1977) Estimating an O-D matrix from network data, a network equilibrium approach. Publication no. 60, Centre de Recherche sur les Transports, Université de Montréal.
- NGUYEN, S. (1984) Estimating origin-destination matrices from observed flows. Publication no. 263, Centre de Recherche sur les Transports, Université de Montréal.
- OECD ROAD RESEARCH GROUP (1974). Urban traffic models: possibilities for simplification. OECD, Paris.
- OH, J. (1989a) Estimation of trip matrices in networks with equilibrium link flows. Proceedings of the 21st UTSG Annual Conference, Napier Polytechnic, Edinburgh. (unpublished)
- OH, J. (1989b) Estimation of trip matrices in networks with equilibrium link flows. Proceedings of the 17th PTRC Summer Annual Meeting, University of Sussex, September, 1989, Seminar C, 65-76.
- OH, J. (1989c) Estimation of trip matrices from traffic counts: an equilibrium approach. IMA Conference on Mathematics in Transport Planning and Control, Cardiff, September, 1989.

- OUTRAM, V.E. and THOMPSON, E. (1978) Drivers perceived cost in route choice. Proceedings of the 6th PTRC Summer Annual Meeting, July 1978, PTRC Education and Research Services Ltd., London.
- QUANDT, R.E. and BAUMOL, W.J. (1966) The abstract mode model: theory and measurements. Journal of Regional Science, 6, 13-26.
- RATCLIFFE, E.P. (1972) A comparison of drives route choice criteria and those used in current assignment processes. Traffic Engineering and Control, 13, 526-529.
- ROBERTSON, D.I. (1984) Estimating origin-destination flows by simulating trip choice. Traffic Engineering and Control, 25(7/8), 358-364.
- ROBILLARD, P. (1975) Estimating the O-D matrix from observed link volumes. Transportation Research, 9(2/3), 123-128.
- ROSE, G., DASKIN, M.S. and KEPPELMAN, F.S. (1988) An examination of convergence error in equilibrium traffic assignment models. Transportation Research, 22B(4), 261-274.
- SHANNON, E.S. (1948) The mathematical theory of communication. Bell Systems Technical Journal, 27, 379-423, 623-656.
- SHEFFI, Y. and POWELL, W. (1981) A comparison of stochastic traffic assignment over congested networks. Transportation Research, 15B(1), 53-64.
- SMITH, M.J. (1979) Existence, uniqueness and stability and traffic equilibria. Transportation Research, 13B(4), 295- 304.
- SMITH, M.J. (1984a) Two alternative definitions of traffic equilibrium. Transportation Research, 18B(1), 63-65.
- SMITH, M.J. (1984b) A descent algorithm for solving a variety of monotone equilibrium problems. Proceedings of the Ninth International Symposium on Transportation and Traffic Theory, Delft, 273-297.

- SMITH, R. and McFARLANE, W. (1978) Examination of a simplified travel demand model. Transportation Engineering Journal of ASCE, 104.
- SMOCK, R. (1962) An iterative assignment approach to capacity restraint on arterial networks. Highway Research Board Bulletin, 347, 60-66.
- SPIESS, H. (1987) A maximum-likelihood model for estimating origin-destination matrices. Transportation Research, 21B(5), 395-412.
- STEER, DAVIES and GLEAVE (1987) Validity and application of matrix estimation techniques in trunk road appraisal: final report. Steer, Davies and Gleave, Ltd., London.
- TAMIN, O.Z (1988) The estimation of transport demand models from traffic counts. PhD thesis, Transport Studies Group, University College London.
- TURNQUIST, M. and GUR, Y.(1979) Estimation of trip tables from observed link volumes. Transportation Research Record, 730, 1-6.
- VAN VLIET, D. (1976) Road assignment I, II and III. Transportation Research, 10(3), 137-157.
- VAN VLIET, D. (1979) The common man's guide to the theory of equilibrium assignment. Technical Note 13, Institute for Transport Studies, University of Leeds.
- VAN VLIET, D. (1987) The SATURN 7 Document. Institute for Transport Studies, University of Leeds.
- VAN VLIET, D. and DOW, P.D. (1979) Capacity-restrained road assignment. Traffic Engineering and Control, 20(6), 296-235.

- VAN VLIET, D. and WILLUMSEN, L.G. (1981) Validation of the ME2 model for estimating trip matrices from traffic counts. Proceedings of the Eighth International Symposium on Transportation and Traffic Theory, Toronto, 640-655.
- VAN ZUYLEN, H.J. (1978) A method to estimate a trip matrix from traffic volume counts. Proceedings of the sixth PTRC Summer Annual Meeting, University of Warwick.
- VAN ZUYLEN, H.J. (1981) Some improvements in the estimation of an O-D matrix from traffic counts. Proceedings of the Eighth International Symposium on Transportation and Traffic Theory, Toronto, 656-671.
- VAN ZUYLEN, H.J. and WILLUMSEN, L.G. (1980) The most likely trip matrix estimated from traffic counts. Transportation Research, 14B(3), 281-293.
- VAN ZUYLEN, H.J. and BRANSTON, D.M. (1982) Consistent link flow estimation from traffic counts. Transportation Research, 16B(6), 473-476.
- WARDROP, J.G. (1952) Some theoretical aspects of road traffic research. Proc. Inst. Civil Engn. Part II, 1(2), 325-378.
- WILLS, M.J. (1978) Linear and non-linear estimator of the matrix. PhD thesis, Department of Geography, University of British Columbia.
- WILLS, M.J. (1986) A flexible gravity-opportunities model for trip distribution. Transportation Research, 20B(2), 89-111.
- WILLUMSEN, L.G. (1978a) Estimation of an O-D matrix from traffic counts: A review. Working Paper no. 99, Institute of Transport Studies, University of Leeds.
- WILLUMSEN, L.G. (1978b) O-D matrices from network data: A comparison of alternative methods for their estimation. Proceedings of the Sixth PTRC Summer Annual Meeting, University of Warwick.

- WILLUMSEN, L.G (1981a) An entropy maximizing model for estimating trip matrices from traffic counts. PhD thesis, Department of Civil Engineering, University of Leeds.
- WILLUMSEN, L.G. (1981b) Simplified transport models based on traffic counts. Transportation, 10(3), 257-278.
- WILLUMSEN, L.G. (1982) Estimation of trip matrices from volume counts, validation of a model under congested conditions. Proc. of the 10th PTRC Summer Annual Meeting, University of Warwick, July 1982. Seminar Q, P230, 311- 323.
- WILLUMSEN, L.G. (1984) Estimating time-dependent trip matrices from traffic counts. Proceedings of the Ninth International Symposium on Transportation and Traffic Theory, Delft, 397-411.
- WILSON, A.G. (1967) A statistical theory of a spatial distribution models. Transportation Research, 1(3), 253-269.
- WILSON, A.G. (1970) Entropy in urban and regional modelling. Pion Ltd, London.
- WOOTTON, H.J., NESS, M.P. and BURTON, R.S. (1981) Improved direction signs and the benefits for road users. Traffic Engineering and Control, 22, 264-268.

APPENDIX 1. LIST OF MAJOR NOTATION

A1.1 Roman letters

A_i	Balancing factor related to trips O_i generated from origin i
B_j	Balancing factor related to trips D_j attracted to destination j
$c_a(V_a)$	Cost on link a when link flow is V_a
C_{ij}	Travel cost between origin i and destination j
\bar{C}_{ij}	Observed travel cost between origin i and destination j
C_{ij}^*	Travel cost in the equilibrium condition between origin i and destination j
C_{ijr}	Route cost used by route r between origin i and destination j
D_j	Total number of trips attracted to Zone j
$f(C_{ij})$	Deterrence function used in the gravity model
$G(\underline{T}, \underline{V})$	Gap penalty function used in the simultaneous method
$G_1(\underline{T}, \underline{V})$	Gap penalty function used in the improved sequential method
g_r	Feasible flows on path r
$\underline{g} = \{g_r\}$	Vector of g_r
h_r	Equilibrium flows on path r
$\underline{h} = \{h_r\}$	Vector of h_r
I	Set of observed link flows
L	Set of links in the network
M	Number of links observed
N	Number of zones in the network
O_i	Total number of trips generated from Zone i
P_{ij}^a	Proportion of trips using link a between origin i and destination j
$\underline{P} = \{P_{ij}^a\}$	Vector of P_{ij}^a
P_{ij}	Set of paths through the network between origin i and destination j
$S_0(\underline{T}, \underline{t})$	Full entropy function
$S_0(\underline{T})$	Full entropy function when $t_{ij}=1$ for all i and j
$S_1(\underline{T}, \underline{t})$	Simplified entropy function
$S_1(\underline{T})$	Simplified entropy function when $t_{ij}=1$ for all i and j
t_{ij}	Prior trips between i and j
$\underline{t} = \{t_{ij}\}$	Prior trip matrix
$t..$	Total number of prior trip matrix

T_{ij}	Estimated trips between origin i and destination j
$\underline{T}=\{T_{ij}\}$	Estimated trip matrix
$T_{..}$	Total number of estimated trip matrix
\hat{T}_{ij}	Observed trips between origin i and destination j
T_{ijr}	Number of estimated trips for route r between origin i and destination j
V_a	Modelled flows on link a
\underline{V}	Vector of V_a
\hat{V}_a	Observed flows on link a
$\hat{\underline{V}}$	Vector of \hat{V}_a
V_a^*	Equilibrium flows on link a
\underline{V}_a^*	Vector of V_a^*
V_{\min}	Minimum link flows on the maximum cost paths

A1.2 Greek letters

α_{aij}	Coefficient related to link a for the i-j pair
β_{aij}	Coefficient related to link a for the i-j pair
β	Cost perception parameter to be calibrated in the gravity model
δ_{ar}	$\delta_{ar}=1$, if link a is used by route r, $\delta_{ar}=0$, otherwise.
δ_{ijr}^a	$\delta_{ijr}^a=1$ if link a is used by route r between i and j, and $\delta_{ijr}^a=0$, otherwise.
δT_{ij}	Perturbed trips between origin i and destination j
$\underline{\delta T}$	Vector of δT_{ij}
ΔT	Absolute value of $\underline{\delta T}$
ϕ	Scale parameter
λ	Optimal linear combination parameter in the Frank-Wolfe equilibrium assignment method
λ_a	Lagrange multiplier related to the flow constraint of link a.
μ	Penalty function parameter
μ_i	Penalty function parameter of the ith sub-problem.
ρ_{ij}	Sampling factor for the i-j pair

APPENDIX 2. NGUYEN'S EXAMPLE NETWORK DATA

C	NA	NB	So	Sc	Cap	WS	da	P	I
C	1	5	5	5	9999	1S	100	1.0	1
C	1	12	5	5	9999	1S	100	1.0	2
C	2	5	5	5	9999	1S	100	1.0	3
C	2	9	5	5	9999	1S	100	1.0	4
	5	6	60	30	500	1S	500	1.0	5
	5	9	50	40	400	1S	500	1.0	6
	6	7	70	35	300	1S	500	1.0	7
	6	10	55	37	330	1S	500	1.0	8
	7	8	70	40	400	1S	500	1.0	9
	7	11	50	40	250	1S	500	1.0	10
	9	10	70	55	500	1S	500	1.0	11
	9	13	55	40	350	1S	1000	1.0	12
	10	11	56	45	500	1S	500	1.0	13
	12	6	70	50	350	1S	500	1.0	14
	12	8	50	40	400	1S	1500	1.0	15
8C	3	5	5	5	9999	1S	100	1.0	16
11C	3	5	5	5	9999	1S	100	1.0	17
11C	4	5	5	5	9999	1S	100	1.0	18
13C	4	5	5	5	9999	1S	100	1.0	19

Note) Coded in the SATURN (SATNET) input format

C: if the following node refers to a zone

NA: the A node for the link

NB: the B node for the link

So: link speed (in kph) under free-flow conditions

Sc: link speed (in kph) at capacity level

Cap: link capacity (pcus/h)

W: one way/two-way indicator

S: if speeds were defined

da: link distance(in meters)

P: power to be used in the link flow-delay curve

I: link index