Traffic Congestion and Road User Charges in Australian Capital Cities

Report

Urban travel behaviour is very complex. Analysts have tried to capture its main features in models that provide estimates of the levels and patterns of traffic on the urban road network. While the models are correspondingly complex, they are still radical simplifications of real urban systems. Their treatment of some aspects of travel behaviour can only be described as rudimentary. Nevertheless, they represent the state of the art in quantitative urban transport analysis, and they provide a valuable framework for thinking about urban policy issues.

bureau of transport and
communications economics

Traffic congestion AND Road user charges in Australian capital cities

REPORT

92

AUSTRALIAN GOVERNMENT PUBLISHING SERVICE, CANBERRA

ISSN 1034–4152
ISBN 0 644 36066 6

P A P E R
50% recycled fibres
50% chlorine free bleached
ZANDERS Mega

Cover design by Jean Penny
Desktop publishing by BTCE Information Services

Printed by the Australian Government Publishing Service

FOREWORD

Urban travel behaviour is very complex. Analysts have tried to capture its main features in models that provide estimates of the levels and patterns of traffic on the urban road network. While the models are correspondingly complex, they are still radical simplifications of real urban systems. Their treatment of some aspects of travel behaviour can only be described as rudimentary. Nevertheless, they represent the state of the art in quantitative urban transport analysis, and they provide a valuable framework for thinking about urban policy issues.

Working Paper 15, *Urban congestion: Modelling traffic patterns, delays and optimal tolls*, published by the BTCE in May 1995, presented preliminary work on the potential costs and benefits of reducing congestion in Melbourne through the imposition of differential charges for road use throughout the city.

The methodology developed by the BTCE in Working Paper 15 has also been applied in the current Report to Adelaide, Brisbane, Perth, and Sydney. It has further been extended to provide for the first time a comparison of results obtained by modelling urban congestion charges on the aggregated basis used by many researchers, and the more detailed network approach employed by the BTCE.

The BTCE acknowledges with gratitude the indispensable assistance of R. J. Nairn and Partners in the use and modification of the TRANSTEP model. Professor Max Neutze and Professor Bill Young provided assistance by reviewing the drafts. Belinda Jackson edited the manuscript.

The work in this report was undertaken by Dr Franzi Poldy and Brett Evill.

Dr Leo Dobes
Research Manager

Bureau of Transport and Communications Economics
Canberra
March 1996

CONTENTS

FIGURES

Page

TABLES

ABSTRACT

A commercial traffic simulation package, developed for engineering purposes, was adapted by the BTCE to estimate external and marginal costs of city-wide traffic. A further modification made it capable of calculating simultaneously-optimised road user charges for all road links in a city individually, and predicting the impact of these charges on trip generation, destination and mode choice, and choice of route. Traffic levels, delay, and average levels of charges are presented on a three kilometre grid for six large Australian cities.

The model was also used to estimate the effects of uniform per-kilometre charges. Various levels of such a charge are compared with the 'optimal' uniform charge suggested by a traditional analysis of aggregate modelling results. The bias in traditional estimates of optimal congestion charges is thus quantified.

CHAPTER 1 INTRODUCTION

With the increase in environmental awareness in a growing Australian population over the last two decades, it is natural that public attention should have focussed on the motor car. Ever more frequently, commentators are highlighting the effects of worsening congestion on the streets of cities in Australia and abroad. Noise, accidents, noxious emissions, and, increasingly in recent years, the contribution to national emissions of greenhouse gases, are attracting specific attention.

Although problems such as congestion, noise, accidents and noxious emissions from urban transport predate the internal combustion engine, many popular commentaries include a call to reduce the use of motor cars. Banning them from the Central Business District (CBD), or limiting the availability of parking, is often proposed as a solution in the cause of 'reclaiming the streets'.

As community acceptance of the price mechanism has increased in the era of microeconomic reform, however, attention has also turned to the issue of charging motorists directly for their use of road infrastructure. Construction of new arterial roads in Sydney and Melbourne is being financed primarily through tolls. But the debate is also moving towards the desirability of charging for the use of all roads (not just to finance construction of new ones) as a means of constraining congestion to economically efficient levels.

Determination of economically efficient charges for road use requires detailed knowledge of the marginal costs—both private and in terms of externalities imposed on other road users—and benefits involved. To date, most analyses have examined such costs only in very aggregated form that provides little guidance to policy makers concerned with reducing different levels of congestion in various parts of a city.

Although the original objective of the BTCE's study of the costs of urban traffic congestion was that it should be an adjunct to its work on the costs of reducing greenhouse emissions in the transport sector, this report in fact presents a number of innovative results. In particular, the methodology developed by the BTCE enables:

- estimation of economically optimal road user charges for each link in the urban road network;
- comparison of the results of estimating road user charges on the aggregated basis employed to date in other analyses, and the detailed network approach adopted by the BTCE;
- separate but comparable estimates for six Australian capital cities;
- estimation of potential revenues generated from road user charges, and the net economic benefits accruing to road users from their imposition.

Although the focus of this report is on urban traffic congestion, the methodology developed by the BTCE can readily be applied to the analysis of costs and benefits associated with other traffic-related urban externalities such as noise or noxious emissions.

Congestion was chosen for initial study primarily because of the relatively high cost that it imposes on the community in comparison with other transport externalities (BTCE 1993). This choice was also logical because traffic congestion affects the level of accidents, noise and noxious emissions, and therefore required prior analysis. Although accidents may have a temporary effect on localised congestion levels, the feedback effect in the opposite direction is relatively much weaker.

Transport analysts suffer increasingly from a lack of current data, often because of the costs of collection. This BTCE study is no exception.

Data sets on traffic flows in the various capitals ranged from those collected in the late 1980s to those collected in 1993. In the case of Sydney, for example, it was not possible to capture the effects of the harbour tunnel. While the tunnel represents only one route, its importance is relatively great in Sydney because of the limited number of alternatives in that area of the city.

It is therefore important to interpret the results presented in subsequent chapters as indicative and exploratory only. The purpose of their presentation is to provide policy makers with information on orders of magnitude and to identify patterns, not to propose the implementation of specific charges.

Public discussion of road user charges will inevitably address the issue of revenue and its use by the implementing government or agency. While the issue is an interesting one from a purely economic point of view, it is ultimately a matter of broader government policy. It has therefore not been discussed in this report.

CHAPTER 2 URBAN TRANSPORT MODELS

The following outline of the general features of urban transport models aims to provide a qualitative understanding of what has been taken into account in reaching the results presented in later chapters.

The description is at a generic level and does not necessarily refer to the TRANSTEP model from which the results were obtained. For a detailed description of TRANSTEP see Nairn, Field & Parker (1977) or the current General Description (Nairn & Partners 1986) or User Manual (Nairn & Partners 1991). In fact, TRANSTEP differs in some respects from the commoner so-called traditional four-step transport models and, where this is significant, it is noted. However, as has been mentioned, the results could, in principle, have been obtained with other models, and it is important for this work that detailed differences should not distract from the essential features of the models and the results.

MODEL COMPONENTS

Land use in the city is represented by land use statistics for a set of zones that cover the urban area. The transport system is described in terms of the road network of nodes and links and public transport routes which use the roads or their own dedicated links. The numbers of zones, nodes, links and other components in the model of a particular city depend on the size of the city, the availability of data and the aims of the analysis.

Zones

The urban area is divided into zones, for each of which land use characteristics relevant to travel demand are recorded. The most important land use characteristics, and the ones generally used in TRANSTEP, are:

- residential—represented either by population or by number of households, usually with an indication of the income distribution;
- employment—represented by numbers of jobs, possibly with some disaggregation by industry type;
- retail activity—represented either by numbers of retail jobs or by area of retail floor space;
- educational facilities—represented by enrolments at schools, colleges and universities; and
- recreational—which may be represented in a variety of ways such as the areas of parkland or the spectator capacity of sporting facilities. This aspect of land use is less important for the analysis of peak hour commuting travel, which is the focus of most urban transport modelling work.

External zones can also be defined to represent travel origins and destinations outside the urban area. These are generally more important for smaller cities where travel with an external end point is a greater proportion of total urban travel than in the larger cities.

The road network

The road network is represented by a set of nodes and links. Nodes represent junctions or intersections of roads, or in some cases serve to indicate the route of roads that are not straight. Nodes are generally numbered and, for plotting purposes, their co-ordinates in the plane are given. A subset of nodes, one for each zone, are known as zone centroids and represent the interface between the zone and the road network.

In some models, nodes have additional attributes representing the junction or intersection type (stop, give way, round-about, traffic signals, etc) and its impact on travel time. TRANSTEP does not model intersection delay, but the consequences of delay are implicitly incorporated in the speed–flow relationships on the road links adjacent to the intersections.

A link is a stretch of road joining two nodes, and is identified by the numbers of the pair of nodes it connects. Links are of a number of types corresponding to the different types of road in the network—freeways, divided or undivided arterials, collectors and distributors—with account taken of number of lanes. For city-wide strategic modelling, the network would typically include only these larger road types. Streets within zones are represented in an approximate way by centroid connectors, which are a particular type of link connecting zone centroids to nodes on the road network proper.

Each road type is characterised by a free speed and a traffic capacity. The free speed is the speed of vehicles at very low levels of traffic. The traffic capacity is a loose measure of the maximum traffic volume (in vehicles per hour) that the road can carry. These quantities are parameters in the speed–flow relationship, which describes how traffic speed falls off as the volume of traffic increases (see chapter 3 and appendix I).

Public transport routes

There is a distinction between on-road and off-road modes. On-road modes such as buses and trams share the road network with private traffic; they thus contribute to and experience the effects of congestion. Off-road modes such as trains and ferries operate between nodes of the network on dedicated links that are not part of the road network. Routes for both on- and off-road modes are defined as consecutive sequences of links between nodes. Nodes on public transport routes can also be classified according to such features as whether passengers can board and alight or whether there are special facilities for transfer between routes.

Each public transport route is characterised by its scheduled speed, frequency of service and capacity (persons per vehicle). In addition, for on-road modes, the contribution of the vehicles to congestion is recorded in terms of an equivalent number of passenger car units,[1] and the actual speed is limited by the traffic speed, which may be less than the scheduled speed in congested conditions.

MODELLING URBAN TRAVEL

Transport models try to describe people's use of the urban transport system to meet their needs for access to all the facilities of the city, and the consequences of this use for traffic on the road and public transport

1. For example, if a bus were equivalent to three cars for congestion purposes, a traffic flow of 1000 vehicles per hour made up of 980 cars and 20 buses would be

networks. People's travel behaviour is extremely complex and depends on (among other factors): where they live; where they work; where they shop or go to school; the number and types of vehicles they own; their access to, and the cost and convenience of, public transport; when they choose to travel and whether they have any choice in the matter; and so on. Many of these factors are, themselves, dependent on transport conditions in the city; so there are important feedback effects on different time scales that need to be taken into account. In the short term, congestion, travel time and transport costs influence people's choice of mode and where to shop. In the longer term, these factors influence where people live and work, and where developers locate new facilities. And the location of population, employment and other facilities influences the roads and transport services provided by governments and the private sector. In principle, the models try to capture all of this.

In practice, the representation of these complexities, in all models, is very uneven. The academic literature describes a great deal of research into urban travel but, until very recently, it has not been feasible to implement many of the findings of this research in models suitable for routine use.

A body of practical understanding and feasible techniques has evolved over the last 30 years, and forms the basis for almost all urban transport models in common use. The standard approach has become known as the 'traditional four-step modelling process'. The following paragraphs provide a brief overview of this process. The TRANSTEP model does not, in fact, fall quite within the four-step tradition. However, for the purposes of this paper, it is the general principles and limitations that are important, and these are common to all models.

The four steps of the traditional process are:

1. trip generation—estimating the total number of trips *to* and *from* each zone, on the basis of people's needs to be elsewhere and of the power of urban facilities such as job locations, schools and shopping centres to attract people;

2. trip distribution—determining the origin and destination zones of the trips actually made;

3. mode split—estimating, on the basis of the travel time, cost and convenience of the trips to be made, the proportion of trips between each pair of zones that will be made by public transport or by private car;

4. trip assignment—estimating the paths taken by traffic on the road network and the public transport routes used to accomplish the required trips.

Trip generation and distribution and the origin–destination matrix (steps 1 and 2)

The aim of the first two steps of the traditional process is to produce a zone-to-zone origin–destination (O–D) matrix of person trips to represent the demand for travel about the city. The O–D matrix is a square matrix, of order equal to the number of zones, in which the entry in row i and column j records the number of trips from zone i to zone j. Clearly, the sum of all the entries in row i is the total number of trips originating in zone i, and depends, in part, on the number of people in the zone who are likely to want to make trips. Similarly, the sum of the entries in column j is the total number of trips that end in zone j, and depends, in part, on the extent of the facilities in the zone that are likely to attract people.

The first step, trip generation, is concerned essentially with generating the row and column totals (whose grand totals must, of course, be equal—the total number of trips *from* all zones must equal the total number of trips *to* all zones). The second step, trip distribution, is concerned with filling in the body of the O–D matrix in a way that is consistent with the row and column totals *and* with some measure of people's perceptions of the generalised cost[2] of trips between each pair of zones.

Clearly, this procedure is *not* the way individuals make their travel decisions, but it can produce an acceptable description of aggregate trip making.

TRANSTEP departs somewhat from the traditional four-step process by combining the first two steps. The result, however, is still a zone-to-zone O–D matrix of person trips which is the input to the mode split routines in the third step. The details are described in Nairn and Partners (1986).

---

2. The notion of generalised cost is used widely in transport economics; not always very precisely. Its main use is to provide an aggregate measure of all the cost components (including non-monetary costs) relevant to a particular analysis. In this work, the focus is on the costs taken into account in making travel decisions (whether to travel, where to travel, by what mode and by what route). The generalised cost includes vehicle operating costs, public transport fares and the value of travel time.

It is worth noting, in passing, that the costs of trips between pairs of zones are stored in a cost matrix of the same order as the O–D matrix (for technical reasons, the cost matrix is often called a 'skim' matrix). The entry in row i and column j of the skim matrix records a measure of the cost of trips from zone i to zone j. Different cost measures are used for different purposes and at different stages of the calculations. For example, a skim matrix might record the distance (on the shortest route through the network) between each pair of zones, or the travel time (at free speed or in congested conditions), or a weighted average of the costs by private car and public transport.

Mode split (step 3)

In the third (mode split) step, the number of trips between each pair of zones (the entries in the O–D matrix) is allocated among the available modes. The modelling in this step is closer to the way individuals make their travel decisions in real life, and attempts to take account explicitly of the main factors thought to govern choice of transport mode: access time; frequency of service; need to transfer between routes; in-vehicle time; comfort; fares, overall travel time and vehicle operating costs.

The principal split sought at this stage is that between private car and public transport, but different choice structures can be accommodated. In some models, the person trip O–D matrix is first split into motorised and non-motorised trips. The proportion of non-motorised trips, representing walking and cycling, naturally falls off rapidly with trip distance. These trips are, therefore, primarily intra-zonal. The motorised trip matrix is then split between car and public transport trips. The entries in the car trip matrix may be divided by an average vehicle occupancy (which may be a function of trip length) so as to represent vehicle trips as required in the traffic assignment step. The public transport trip matrix continues to represent person trips.

Assignment (step 4)

The final step is to determine how all these trips will be accommodated by the road network and the public transport system. The routes of trips by both public transport and private car are selected to minimise their generalised cost, which can include fares and vehicle operating costs, but which is often dominated by travel time. For low traffic volumes, the preferred route for a trip is close to the shortest path through the network between origin and destination. But, as traffic volumes increase and congestion reduces speeds on preferred routes, trips may be diverted to routes of greater distance but shorter travel time.

The outcome of this fourth step, the so-called equilibrium assignment, is a pattern of traffic on the road network (and patronage of public transport services) such that all trips are made at minimum generalised cost. The assignment is said to be in equilibrium because no trip could be made by an alternative route at a lower generalised cost.

These patterns of traffic and public transport patronage are the principal outputs of urban transport models. Depending on requirements, traffic volumes, speeds and the state of congestion on every link of the road network can be determined. Chapter 5 describes how these link volumes are used to estimate the costs of congestion, the level of optimal congestion road user charges and the revenue and net benefit from imposing such charges.

Computation, feedback and iteration

Nothing has been said above about the details of the calculations. For the most part they are not necessary for an overview. However, one aspect of the calculations does need to be noted.

Traffic gives rise to congestion and delay. But delay (acting through travel time and the generalised cost of trips) is one of the factors that influence trip demand and mode choice, which together determine traffic volumes. The extent of this feedback is not known at the start of the model calculations, and it is therefore necessary to iterate through the whole process a number of times until consistent results are obtained.

In the first iteration, travel demand (as represented by the O–D matrix of person trips) and mode choice are determined on the assumption that trips will be made in traffic flowing at its free speed. Traffic assignment on this basis takes no account of congestion and, as a result, unreasonably high traffic volumes are assigned to some links. These high volumes imply very large delays. This reduces the demand for travel on these links in the second iteration, at the end of which improved estimates of traffic volumes and travel times are obtained. It is *not* obvious that this procedure must necessarily converge, and it is therefore common to use weighted averages of the travel times from the last two iterations as input to the next.

DATA AND CALIBRATION

Urban transport models contain large numbers of parameters. The basic behavioural parameters describing the propensity to travel and mode preferences are estimated from data obtained from detailed household travel surveys. Such surveys are expensive and are conducted only at infrequent intervals.

Land use information (population, employment and retail, educational and recreational facilities in each of the zones) is obtained from census data, supplemented by a variety of other sources.

Information on the road network and public transport services may usually be obtained from the relevant State authorities who are also the source of most of the data required for calibration.

A number of technical parameters are obtained from the academic and applied literature on traffic engineering and management. These include such items as: road capacities and free speeds on different types of road; speed–flow relationships, which describe the impact of congestion; and equivalence factors relating trucks and public transport vehicles to cars for the purpose of estimating congestion. There is considerable uncertainty about some of these parameters.

The models are calibrated by adjusting the parameters that specify travel behaviour until the model descriptions of urban travel patterns match reality as closely as possible. Calibration data are often difficult to obtain. Household travel surveys provide information on trip length distributions and mode splits. Traffic counts on the main arterial roads are generally available, although care is required to ensure that the period to which the data refer corresponds to the period modelled. In some cases, screen line data are available. Screen line data record the number of vehicles on all roads crossing a specified line during a specified period. Public transport authorities can provide additional mode split information, although, again, it is important to ensure that it refers to the period modelled.

Economic interpretations of the model results are often dominated by the value attributed to changes in travel time. Model predictions of traffic flow speeds would therefore seem to be important items for calibration. Unfortunately, useable speed data for this purpose are rarely available.

LIMITATIONS OF URBAN TRANSPORT MODELS

There are three main weaknesses in the current generation of urban transport models.

The treatment of travel demand

Urban transport models possess all the obvious weaknesses associated with trying to construct useable simplifications of very complex systems. Broadly speaking, the treatment of travel demand (the first two of the four steps of the traditional process) is weak, while the treatment of mode split and traffic assignment is better. While there are simplifications and possibly errors in all the steps, the treatment of travel demand in the first two steps is too aggregate to capture the behavioural richness of individual choice that governs urban travel patterns in real life.

This limitation of traditional transport models is being addressed in another BTCE project in which the Institute of Transport Studies (ITS) at the University of Sydney is developing, under contract, a much more sophisticated model of urban passenger travel demand Hensher (1993). Based on a unique household travel survey that covers all the major Australian capital cities, the model will include an explicit treatment of the choice of: residential location; dwelling type; workplace location; work patterns (as they affect commuting); number and type of vehicles per household; vehicle use for different purposes; mode for the journey to work; and commuting departure time.

The BTCE/ ITS model represents a very substantial improvement in the treatment of urban travel demand. Pending its incorporation in routine urban transport models, the emphasis in this report is on those aspects of the current models that depend most strongly on step 4, the traffic assignment step.

Omission of freight transport

The current generation of urban transport models deal only with passenger travel. This is clearly an important limitation, and a variety of tactics have been adopted to get round it—if only partially. The simplest procedure is to calibrate the models against actual traffic volumes on the assumption that some fixed proportion (typically 5 to 15 per cent) is composed of freight vehicles. The further, usually implicit, assumption here is that freight and passenger traffic follow the same patterns. This is clearly wrong.

A slightly better procedure, if some information about freight traffic flows is available, is to pre-load the freight traffic onto the road network, and to conduct the analysis of passenger traffic flows against a constant background level of freight traffic. In this way, freight contributes to congestion and the other impacts of traffic, although it does not respond endogenously to changes within the model. The impact of freight policies can, however, be assessed by varying the freight traffic flows exogenously.

Mainly because of a lack of behavioural data, there appears to be no prospect, in the foreseeable future, of including a treatment of freight transport decision making in the models in the way that is done for passengers.

Focus on peak hour commuting travel

Urban transport models arose to meet the need for analysis to support investment decisions relating to urban transport capacity. As the demands on that capacity were greatest during peak hour commuting travel, it was natural that modelling should focus on this period and on the journey to work. With further growth in the demand for urban travel the inter-peak periods have filled in to some extent (though with different patterns of traffic flow). As a result, the peaks are not now so clearly distinguishable. Nor is the journey to work as dominant a proportion of overall travel.

It is possible and, indeed, fairly common to set up the models to represent all-day travel. The problem with this is that the model parameters are required to represent averages over even wider ranges of behaviour than is the case with peak hour modelling. Congestion phenomena, in particular, occur on a short time scale (where even the typical one-hour modelling period may be too long) and cannot be treated adequately in an all-day model.

When there is a requirement for estimates of aggregate costs or impacts on a whole-day or annual basis it would, in principle, be necessary to set up models specifically for each of the characteristic time periods and traffic patterns. Analytical resources and data usually do not permit this. The alternative is to scale the results from peak hour analyses, taking account of their special features as far as possible. There are obvious limits to the accuracy of such procedures.

CHAPTER 3 TRAFFIC OPERATIONS ON ROADS AND NETWORKS—THE SPEED–FLOW RELATIONSHIP

Time costs dominate the economic analysis of congestion and determine the levels and patterns of traffic on urban road networks. Analyses of congestion therefore depend ultimately on assumptions about the speed–flow relationships which describe how traffic speeds are reduced as traffic volumes increase.

This is the domain of traffic engineering. A great deal is known at a detailed level, but it has proved difficult to generalise this knowledge in a way that makes it suitable for economic analysis and whole-city strategic modelling. Strategic analysis is therefore based on ad hoc relationships which reflect traffic engineering knowledge in a qualitative way.

SIMPLE THEORY OF TRAFFIC FLOW

The simple theory of *steady state*, uninterrupted traffic flow is the usual starting point for derivations of speed–flow relationships and travel time functions. The theory deals with an idealised situation (most nearly realised on freeways, well away from entry and exit points), in which the only constraints on vehicles are their interactions with other vehicles in the traffic stream.

The theory is concerned with three basic quantities: traffic speed, traffic density and traffic flow.

Traffic speed (S) is the average speed of vehicles in the traffic stream, measured in kilometres per hour (km / hr).

Traffic density (D) is a function of the spacing between vehicles, and is measured in vehicles per kilometre (veh / km).

Traffic flow (F), *in the traffic engineering sense*, is the rate at which vehicles pass a reference point, and is measured in vehicles per hour (veh / hr).

The fundamental relationship is shown in figure 3.1. At very low densities, traffic travels at its free speed S_{max}. Speed declines as density increases and becomes zero when density reaches its maximum value D_{max}.

This relationship between speed and density is fundamental, because traffic density is apparent to drivers and governs their choice of speed. The quantity which appears to be of greater economic interest, however, is the traffic flow. This can be derived from speed and density through the relationship

$$F = S \times D$$

The flow is therefore represented by the area of the rectangle between the origin and any point on the speed–density curve. Clearly the area of this rectangle (and the traffic flow) is zero when $D = 0$ and $S = S_{max}$, and when $D = D_{max}$ and $S = 0$. Between these points, the traffic flow rises to a maximum F_{max}, which can be considered the capacity of the road.

Figure 3.2 shows the relationship between speed and traffic flow. The upper branch of the curve shows that speed decreases from S_{max} as traffic flow increases up to the capacity of the road, F_{max}. Attempts to increase the flow beyond F_{max} result in increased density but *reduced* flow, as shown on the lower branch of the curve. Any feasible traffic flow can therefore be associated with two speeds.

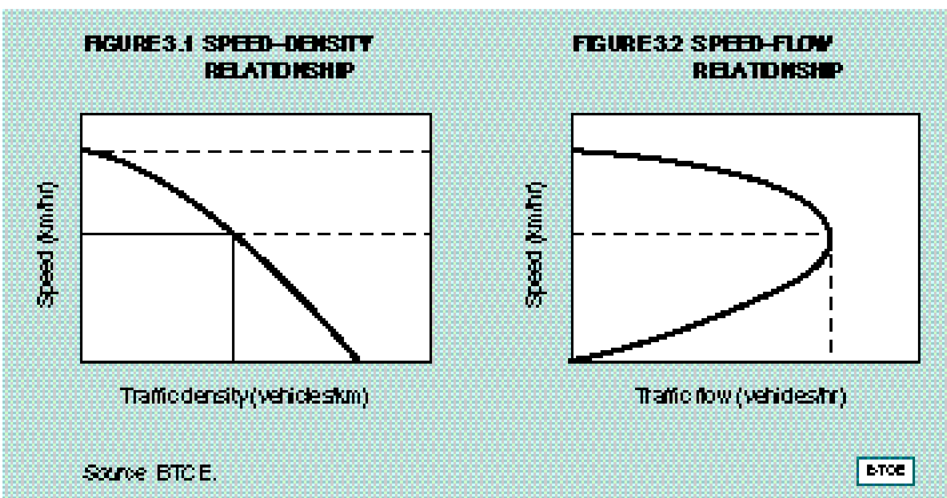


FIGURE 3.1 SPEED–DENSITY RELATIONSHIP

FIGURE 3.2 SPEED–FLOW RELATIONSHIP

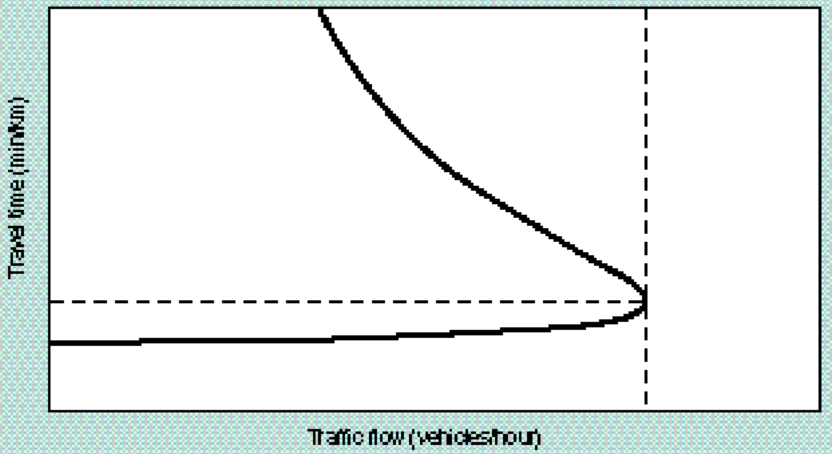
Source: BTCE.

The speed–flow relationship is important for the economic analysis of congestion, because of the costs associated with increases in travel time. By taking the inverse of speed, travel time as a function of traffic flow can be obtained. This is shown in figure 3.3. If travel time per kilometre is multiplied by the value of time, and the cost per kilometre of vehicle operations is added, a curve for the average variable cost of travel is obtained.

There is a certain amount of confusion and controversy about the lower branch of the curve in figure 3.2 and the corresponding upper branch in figure 3.3. Some have denied that travel represented by points on these branches occurs in practice. Others acknowledge its occurrence, but deny its significance for economic analysis on the grounds either that the same traffic flow could be obtained more cheaply on the other branch or that the true economic demand is for complete trips rather than some level of traffic flow (Hills 1993). Others again have taken the economic significance of such travel seriously, and have examined its stability under different demand assumptions (Hau 1992). Finally, it has even been suggested that such travel can, in certain circumstances, be socially optimal (Else 1981, 1982, 1986).

Much of the confusion seems to be due to two related limitations of the analysis. The first is that it may not be possible to capture the essential issues in a purely steady state analysis. The second is that the traffic engineering concept of traffic flow is not strictly what is required for economic analysis.

FIGURE 3.3 TRAVEL TIME FUNCTION



Source BTCE.

From a pragmatic modelling point of view, as well, the travel time function in figure 3.3 is inconvenient. As described in chapter 2, travel time functions are required as part of the traffic assignment step of the modelling process. For each link of the network it must be possible to determine a single travel time for any proposed traffic flow. But the modelling process has no information about traffic density to permit selection between the two branches. More importantly, the function must provide a travel time for *any* level of traffic flow, including flows greater than F_{max}. What is required is a *single valued* function for *all* values of flow.

QUEUING DELAYS

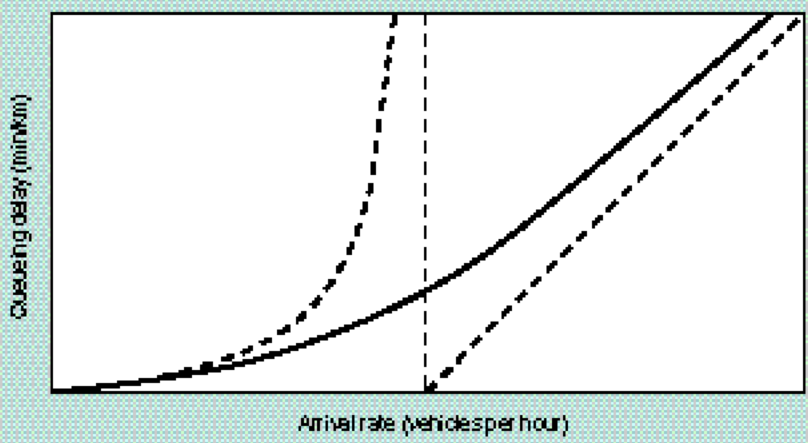
The simple steady state scheme shows that traffic flow *on the road* cannot exceed F_{max}. But it is certainly possible that the rate of arrival of vehicles wishing to use the road might exceed F_{max}. During such a period, the vehicles in excess of F_{max} will not be accommodated on the road—*they will have to wait*. Obviously, if this situation persists (as it must in the steady state), the queue of waiting vehicles would grow indefinitely (contradicting the steady state assumption). In fact, of course, the rate of arrivals eventually falls below F_{max}, and the queue dissipates. The vehicles which made up the queue will, of course, have suffered a queuing delay in *addition* to the time spent on the road, and this additional delay should be included in the total travel time.

This suggests that the situation described in the last section and illustrated in figures 3.2 and 3.3 is *not the whole story*. In terms of queuing theory, it describes a service facility (the road) and the customers (the vehicles) currently being served, but it does not include the queue of those waiting to be served, or the time they spend in the queue. It also suggests that a purely steady state treatment will not capture the essentials of the congestion problem.

It will also be necessary to distinguish the simple traffic engineering notion of traffic flow (the rate of flow of vehicles past a point) from the more economically relevant concept referred to above as 'the rate of arrival of vehicles wishing to use the road'.

Figure 3.4 illustrates some elementary results from queuing theory which are needed to complete the story. The figure shows how average delay varies with arrival rate for a queuing system with a (maximum) service rate F_{max}. The dashed lines illustrate two different simplifying assumptions which lead to well-known results. The solid line shows the results for the more complex intermediate region.

FIGURE 3.4 QUEUEING DELAY FOR UNDER- AND OVER-CAPACITY FLOW



Source: Adapted from Kimber & Hollis (1979).

The first case in figure 3.4 (to the left of F_{max}) shows a *steady state* queuing system with *random* arrivals at a fixed average rate less than the service capacity. The well-known result, shown by the dashed curve, is that the average delay tends to infinity as the average arrival rate approaches capacity. The steady state assumption means that arrival rates greater than capacity cannot be considered.

If the steady state assumption is abandoned, the situation becomes more complex. The second case in the figure (to the right of F_{max}) deals with *regular* arrivals at a rate greater than capacity. As the average delays in this situation increase without limit, it is necessary to restrict attention to the average delay suffered by arrivals *during a fixed period of time*. (This is not unrealistic; urban congestion does not last indefinitely, but occurs in relatively well-defined peaks.) The average delay to arrivals during such a period is proportional to the excess of the arrival rate over capacity and to the length of the period. This is shown by the dashed line, drawn for a particular choice of period. For longer periods, the slope of the line would be steeper.

These simple cases give adequate approximations for the average delay when the arrival rates are not near capacity. Both fail near capacity. Kimber and Hollis (1979) derived a function which provides a smooth transition between the two regimes. Their function describes the average delay experienced by *random* arrivals *during a fixed period* for the complete range of arrival rates. This is the solid curve in figure 3.4.

This curve has the desired characteristics of a travel time function for use in traffic assignment. In the spirit of the argument above, these delays would be added to the lower branch of the curve in figure 3.3 to produce a total travel time curve for all values of the arrival rate[3] In fact, the limited knowledge of the parameters of either of the functions and the general complexity of urban traffic operations would not justify this refinement. As the Kimber and Hollis curve by itself has the desired characteristics, it can be used for modelling and analysis without the unnecessary complications of the curve in figure 3.3.

There are, moreover, other reasons for preferring the Kimber and Hollis function.

INTERSECTION DELAY

While the theory of steady state uninterrupted traffic flow is generally invoked as the basis for travel time functions (and, ultimately, average cost curves), these conditions are rarely found in urban travel, and never as a contributor to congestion. The principal cause of urban congestion is intersection delay.

Analysis of queue lengths and delays at road junctions under time-varying traffic demand was, in fact, the context of the work by Kimber and Hollis (1979), though the techniques they used can be applied to any queuing situation. Akçelik (1991) has used these techniques in the Australian context to develop travel time functions and speed–flow relationships for a number of different types of urban roads. The essence of the approach is that the travel time on a link is simply the free speed travel time plus the average delay associated with queuing at obstructions, of which intersections will usually be the most important. The average queuing delay is estimated using functions of the form shown in figure 3.4 so that finite periods of overcapacity flow are taken into account.

3. This is not to suggest that there is anything unreal about the upper branch of the curve in figure 3.3. Episodes of low speed due to high traffic density are common, and clearly influence traffic flows. In fact, they are one of the mechanisms which contribute to randomness in traffic operations and, strictly speaking, should have been taken into account in the queuing discussion above because they result in the average service rate of the road system being less than F_{max}.

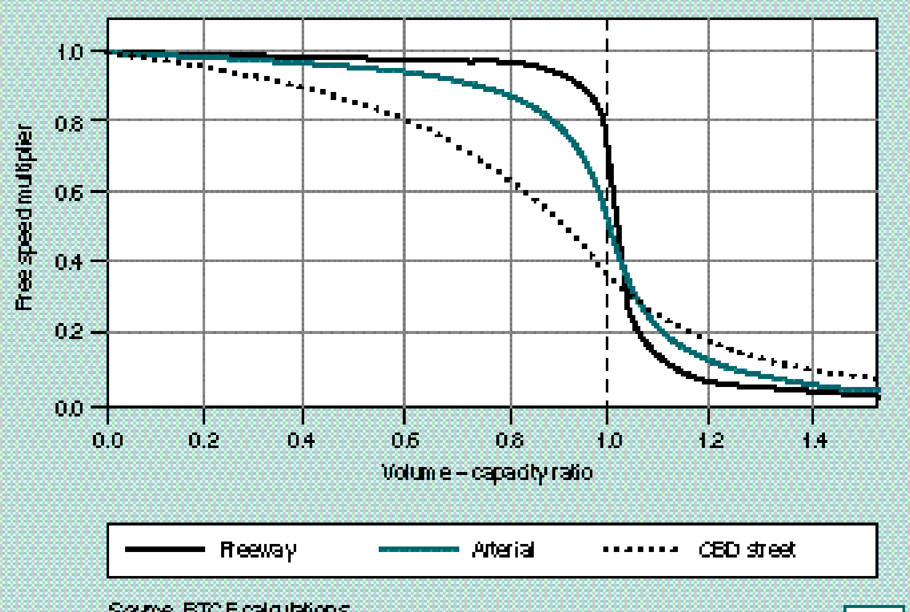
SPEED–FLOW RELATIONSHIPS

The speed–flow relationships proposed by Akçelik are shown in appendix I together with those adopted for this study. The relationships used in this report have the same functional form as Akçelik's, though with somewhat different parameter values.

Figure 3.5 shows the speed–flow relationships for three of the road types: freeways, arterials and CBD streets. In order to emphasise the different shapes of the curves, they are presented in a normalised form in which speed is measured as a fraction of free speed, and traffic flow as the volume – capacity ratio. The actual free speeds and capacities are, of course, very different for these road types.

The progressive change in shape of these curves as one moves from freeways to CBD streets is characteristic of the increasing contribution of delay at intersections and other obstructions in the overall travel time function. This is described in more detail in appendix I.

FIGURE 3.5 NORMALISED SPEED–FLOW RELATIONSHIPS FOR SELECTED ROAD TYPES



Source BTCE calculations.

The important general point to note, however, is that the relationship is single valued—the curve does not bend back towards the origin. The speed falls to low values near the nominal capacity of the road, and tends to zero as the flow increases. This is not compatible with the traffic engineering definition of traffic flow as the product of speed and density ($F = S \times D$). Rather, it shows that the flow plotted on the horizontal axis is the rate of arrival of vehicles seeking to use the road, and that the speed is an average speed based on distance and travel time—including queuing and other delays. It follows that analyses based on these speed–flow relationships take account of the costs of trips which start *but may not be completed* within the analysis period.

CHAPTER 4 THE ECONOMICS OF CONGESTION

The standard theory underlying the economics of congestion is well known, and has been discussed in the literature for many decades (Pigou 1920). It is not presented in detail here. Small (1992) gives a simple introduction and Hau (1992) provides a very useful diagrammatic analysis. The aim of this chapter is to define and clearly identify the essential economic concepts and the quantities which are of policy interest and for which estimates are to be provided.

The treatment is at a general level and does not cover a number of important issues which arise in actual application of the analysis. These have to do with the interpretation of the demand and cost curves and the units in which the quantity of travel should be measured. They are discussed in chapter 5.

THEORETICAL OVERVIEW

Economic theory attributes the problem of urban traffic congestion to the fact that road users do not take account of the full costs of their travel decisions. If road users did take account of the full costs, they would find that some of their trips were not worth making—in other words, that the benefits of these trips were less than their full costs. From the community's point of view, these low value trips should be rescheduled to less congested periods, or combined with other trips, or even suppressed altogether.

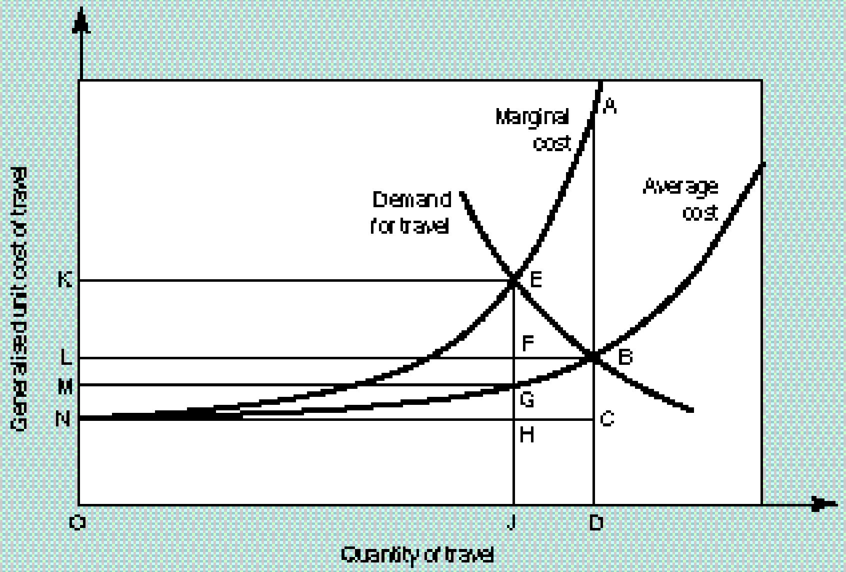
Congestion is a classic example of an economic externality. It arises because some of the costs of the decision to travel are not borne by the decision maker. Road users considering whether to join a congested traffic stream would normally take account of the travel time and vehicle operating costs they would expect to incur. These are the private costs against which they would weigh the benefits and, beyond a certain point, decide not to travel. But road users do not take account of the fact that their decisions to travel increase congestion and impose *additional* costs on others.

Analysis of this situation suggests that overall social welfare would be increased if road users did take account of the full costs of their decisions, and that a suitable road user charge would provide the incentive for them to do so.

The main features of the analysis are most readily understood with the help of the diagram which has become standard and which is reproduced in figure 4.1.

The vertical axis represents the generalised unit cost (or price) of travel. The generalised unit cost is an extension of the concept of cost to include items such as travel time which influence travel behaviour, but which are not usually thought of in monetary terms. It is a prime assumption of this approach to the analysis of congestion that conventional cost components such as vehicle maintenance and fuel *can* be combined with travel time in a generalised unit cost. Naturally, this requires a conversion factor between time and dollars; that is, a value of time.

FIGURE 4.1 ECONOMIC THEORY OF CONGESTION CHARGES



Note The cost of current congestion is area LBCN; the revenue from congestion charges is area KEGM; the net social benefit is area ABE.

Source BTCE calculations.

The horizontal axis represents the quantity of travel. The appropriate units in which this should be measured depend on how the theory is to be applied, which is the subject of chapter 5. For the purposes of this chapter it is not necessary to define this quantity more precisely.

In the diagram, the average cost curve represents the unit cost of travel as perceived by individual road users, and is the basis for individual (private) decision making. It is made up of vehicle operating costs (maintenance and fuel) and travel time costs. It is an average cost, and its product with the corresponding quantity of travel gives the total cost incurred by all road users. The marginal cost curve is the derivative of this total cost with respect to the quantity of travel or, in other words, the contribution to the total cost of the marginal unit of travel. The vertical distance between the marginal and average cost curves therefore represents the additional costs imposed on others, but not taken into account by the marginal user—that is, the *external* costs.

The central point in the analysis is that the marginal traveller's decision to travel is based on private costs (represented by the average cost curve), which are less than the resulting actual increment in total costs (represented by the marginal cost curve).

The demand curve, as usual, represents the benefit of the marginal unit of travel.

Summary of the main features of the analysis

With these definitions, the essential features of the analysis can be summarised and identified with elements of figure 4.1.

The current quantity of travel and its costs

The current quantity of travel OD is determined by the intersection of the demand and average cost curves. It results from decisions by road users who take into account only their own private costs.

The current private costs are equal to the average cost DB of this quantity of travel.

Road users do not take account of the additional costs that their decisions impose on other road users. These are the marginal external costs, given by the difference BA between the marginal and average cost curves.

The socially optimal quantity of travel

The socially optimal quantity of travel OJ is determined by the intersection of the demand and marginal cost curves. This quantity is optimal because it avoids travel beyond the point of intersection of these curves for which the full social costs are greater than the benefits.

The optimal or efficient congestion road user charge

The efficient road user charge required to limit travel to the socially optimal quantity OJ is equal to the marginal external cost at that quantity of travel. It is given by the difference GE between the marginal and average cost curves.

The toll revenue

The revenue raised by the road user charge GE imposed on a quantity of travel OJ is given by the area KEGM. This revenue is a transfer payment from road users to society. It does not change total welfare—though it does change the welfare of the road users who pay the charge.

The current cost of congestion

The increase in the unit cost of travel due to the current level of congestion (over the unit cost in a hypothetical uncongested situation) is represented by NL (or CB). The total cost of congestion to current road users is the product of this increase with the current quantity of travel OD. This total cost is represented by the area LBCN.

The cost of congestion at the social optimum

The cost of congestion in the socially optimal situation is the excess NM (or HG) in the unit cost of travel (over that in the uncongested situation) multiplied by the optimal quantity of travel OJ. The cost of the socially optimal level of congestion is, therefore, given by the area MGHN.

Road users' gains and losses

When faced with a congestion charge, some road users (those whose travel is represented by the range JD) refuse to pay the charge and leave the system—the increased cost of their travel now outweighs its benefits. These users are obviously worse off. The extent of their loss is the consumer surplus they previously obtained from their travel. It is given by the area BEF.

Road users who remain experience reduced congestion, so that they travel at higher speeds and suffer less delay. Their unit cost of travel is reduced by the amount ML (or GF) from OL to OM. The total value to these users of the improved traffic conditions is, therefore, represented by the area LFGM.

However, the road users who remain are the ones who pay the congestion charge GE—and this is greater than the reduction GF in their unit cost of travel. As a group, these users are therefore worse off[4] by the difference between the revenue (KEGM) and the benefits (LFGM) they receive in the form of improved traffic conditions. This loss is represented by the area KEFL.

The net benefit from imposing road user charges

The net benefit gained by society from the introduction of optimal road user charges can be represented in the diagram in three different but equivalent ways. The fact that they are equivalent is, of course, a trivial consequence of the geometry of the average and marginal cost curves. Nevertheless, it is useful to be able to consider the issues from different points of view.

(i) The problem with the current congested situation is that it includes a quantity of travel (in the range JD) for which the total costs exceed the benefits. The net loss on this travel is represented by the area ABE between the marginal cost and demand curves. The avoidance of this loss is the net benefit provided by congestion charges, and is therefore represented by the same area ABE.

(ii) The net benefit from congestion charges can also be built up by carefully considering the gains and losses of the different groups of road users. Road users who refuse to pay the charge and leave the system lose the consumer surplus (BEF) they previously obtained from their travel. Road users who remain and pay the charge benefit by a reduction in total travel costs (LFGM). These are the only contributions to the net change in welfare. The charge revenue is a transfer which redistributes but does not change net welfare. The net benefit can therefore be represented by the area LFGM less the area BEF.

4. The discussion assumes all road users have the same value of time. If this is not the case, users whose value of time is sufficiently above the average may, of course, be better off. But road users *as a group* will still be worse off.

(iii) Evans (1992) considers the output of the road system to be the sum of road users' consumer surplus and the revenue from any congestion charges. Starting from the current congested situation with no charges, introduction of charges leads to an increase in revenue and a reduction in road users' consumer surplus. The increase (from nothing) in the revenue is represented by the area KEGM. The loss in consumer surplus is represented by the area KEBL. The net benefit can therefore be represented by the area KEGM less the area KEBL.

THREE MEASURES OF CONGESTION

Within this analytical framework, three measures can be identified which are likely to be of interest to those considering options for dealing with congestion and, in particular, the possibility of introducing congestion charges. These are:

- the current cost of congestion (area LBCN);
- the net benefits achievable with congestion charges (area ABE); and
- the revenues (KEGM).

The ratios of these measures capture some of what is at stake when questions arise about the equity of congestion charges and whether the proposed policies are appropriately matched to the scale of the problem.

The cost of congestion

The cost of congestion is a simple measure which is sometimes used to highlight the scale of the congestion problem. It is defined as the value of the excess travel time and other resource costs incurred by the current traffic over those that would have been incurred if the current traffic volumes had been able to operate with unit costs characteristic of uncongested free flow conditions.

There are two things to note about this definition. The first is that it refers to an unrealisable hypothetical situation. Current traffic volumes could not actually operate on the existing road system under free flow conditions. The cost of congestion is, therefore, primarily a measure of the scale of the problem, useful in motivating the community and governments to address the issues, but not a measure of the savings to be made. Any actual response to congestion will reduce this cost. It will not eliminate it.

The second thing to note is that the cost is measured with respect to a clearly defined and readily understandable (if unattainable) state of zero congestion. This is an important aid to clear thinking and consistent analysis. But other base lines could be defined. There have been attempts to define congestion as only that degradation of travel conditions beyond some ‘acceptable’ or ‘tolerable’ level which should not itself be considered congested. Such attempts suffer from the arbitrary nature and the difficulty of defining ‘tolerable’ traffic conditions.

Some have been concerned that, because the cost of congestion is defined with respect to an unattainable base line, the socially optimal level of congestion is not zero. This concern may be for the political acceptability of policies such as congestion charging which address the problem but which could be portrayed as failing to solve it. There may be a genuine difficulty here but, if so, it is inherent in the situation.

The net benefit from congestion charges—the ‘cost of doing nothing about congestion’

The net benefit from congestion charges is the measure of what can, in principle, be achieved by tackling the congestion problem. It is of greater policy relevance than the cost of congestion, because there is a serious possibility of obtaining these benefits. Not to attempt to do so is an opportunity forgone and can be described as the ‘cost of doing nothing about congestion’.

The revenue from road user charges

In the framework of welfare economics, the revenue from congestion charges is a transfer payment which does not affect net social welfare. As such, it is neither a cost nor a benefit.

But from the separate points of view of road users and governments the level of revenue is crucially important. The appropriate use of these revenues is an interesting and controversial question beyond the scope of this report.

INDICES OF THE ACCEPTABILITY AND APPROPRIATENESS OF TOLLS

One of the principal objections to congestion charges is that the overall net benefit is obtained at the expense of a transfer of welfare *away* from those who bear the cost of current congestion. The strength of this objection depends primarily on the relative magnitudes of the transfer and the net benefit. Evans (1992) notes that it is unlikely to be politically acceptable to strain after a small net benefit for society as a whole at the expense of a large transfer of welfare away from road users.

The ratio of net benefit to total revenue can therefore be taken as an index of the acceptability of congestion charges.

One would also hope that the scale of a policy response would be matched to the scale of the problem. In this sense, the ratio of toll revenue to the cost of congestion might be an index of the appropriateness of congestion charges.

FACTORS AFFECTING THE INDICES AND MEASURES OF CONGESTION

The effect on estimates of these indices and measures of congestion of different assumptions about the elasticity of demand for travel can be seen by inspecting figure 4.1. The current quantity of travel determines the point B on the average cost curve. Different assumptions about the elasticity of demand would therefore be represented by movement of the point E along the marginal cost curve.

With inelastic demand (demand schedule closer to vertical) the net benefit from congestion charges (area ABE) would be smaller and the revenue (area KEGM) greater.

With elastic demand (demand schedule closer to horizontal) the net benefit would be greater and the revenue smaller.

Demand elasticity assumptions do not affect estimates of the current cost of congestion (area LBCN).

Determining the effect of the form of the average and marginal cost curves is more complicated. Figure 4.2 shows the normalised average and marginal travel time functions[5] for freeways and CBD streets derived from the speed–flow relationships shown in figure 3.5. As time is the major part of congested travel cost, the corresponding cost curves would have essentially the same form. Clearly, determining the relationships among the measures of congestion for the various combinations of

5. The normalised travel time functions are defined analogously to the normalised speed–flow relationships in chapter 3. Time is measured in multiples of the free speed travel time and traffic flow in terms of the volume – capacity ratio.

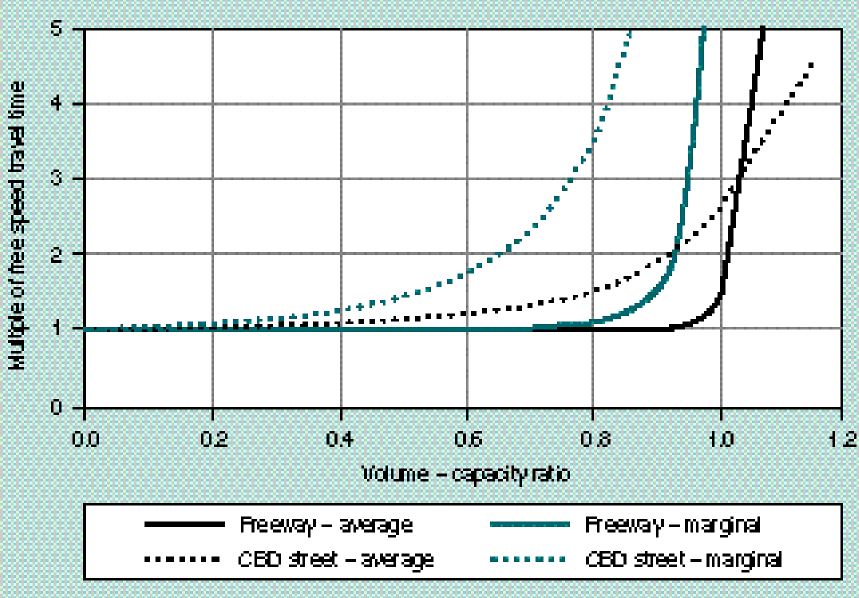
elasticity, degree of congestion and form of the cost curves can become quite involved.

By way of example, figure 4.3 shows how the ratio of net benefit to revenue varies under the influence of these factors. Figure 4.4 shows the corresponding variation of the ratio of revenue to cost of congestion.

These figures provide estimates of the impact and merit of congestion charges under different conditions. The two road types illustrated have been chosen to represent the extremes. The intermediate cases also exist, and urban travel takes place across the whole range of conditions. Assessing the merits of congestion charges for a whole city would therefore need to take account of the contributions of travel under different circumstances.

Alternatively, one could attempt to devise cost functions representative of aggregate urban travel. The pros and cons of these two approaches are discussed in the next chapter.

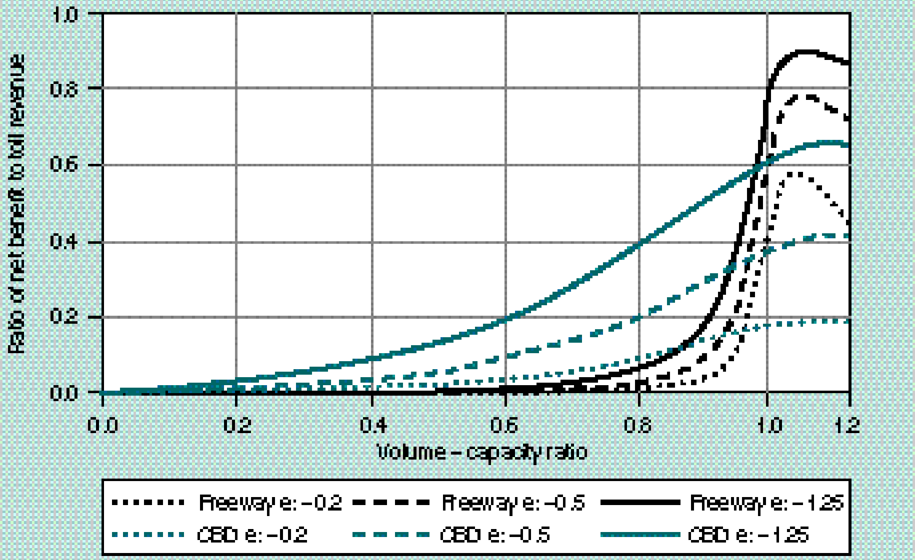
FIGURE 4.2 NORMALISED TRAVEL TIME FUNCTIONS FOR FREEWAYS AND CBD STREETS – INVERSE OF SPEED-FLOW RELATIONSHIPS IN FIGURE 3.5



Source BTCE calculations.

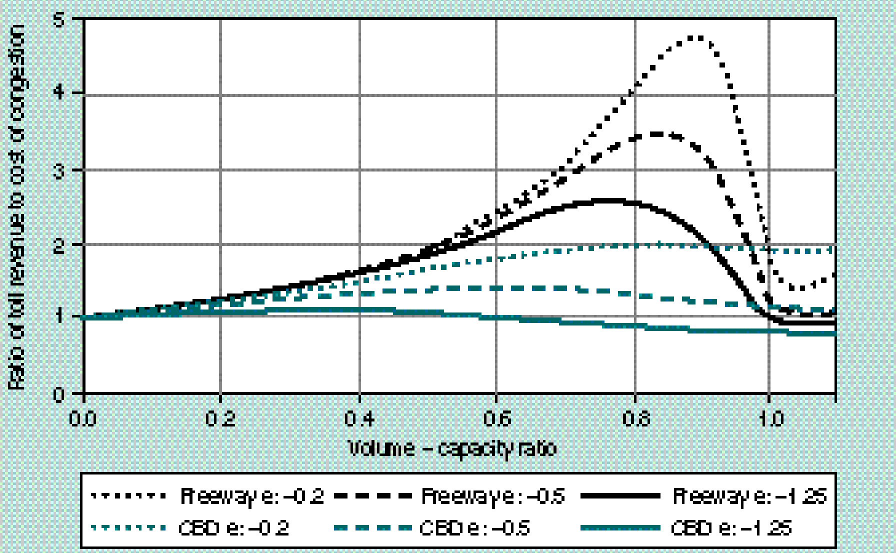
BTCE

FIGURE 4.3 VARIATION OF RATIO OF NET BENEFIT TO TOLL REVENUE WITH ROAD TYPE, DEGREE OF CONGESTION AND DEMAND ELASTICITY



Source BTCE calculations.

FIGURE 4.4 VARIATION OF RATIO OF REVENUE TO COST OF CONGESTION WITH ROAD TYPE, DEGREE OF CONGESTION AND DEMAND ELASTICITY



Source BTCE calculations.

CHAPTER 5 APPLYING THE THEORY

The theory outlined in chapter 4 provides a powerful conceptual framework for addressing the problems of urban congestion. Practical application of the theory, however, is not easy. As mentioned in chapter 4, the theory has been well known for a long time, but attempts to obtain quantitative estimates of congestion charges, revenues and net benefits are much more recent. The main barriers have been the complexity of both urban travel behaviour and traffic operations on urban road networks—and the lack of data.

Most studies bypass the complexities by applying the theory to aggregate travel (measured as total vehicle-kilometres) on the urban road network. This is certainly the simplest approach and the one for which data are most likely to be available. By implication, it is assumed that congestion is a uniform city-wide phenomenon or, at least, that it is adequately described in terms of average measures of traffic conditions (such as speed) over the whole city. Users of this approach would generally acknowledge that it involves gross simplifications, but claim that its results are adequate for broad strategic purposes. They might also claim that the additional complexity of an alternative approach would not be justified by the improvement in the results.

This report is based on the view that the complexities of urban travel behaviour and traffic on networks are important and, in order to take them into account, applies the theory to the traffic flow on individual links of the network as estimated by the TRANSTEP urban transport model.

APPLYING THE THEORY TO AGGREGATE TRAVEL

There are three main difficulties with the aggregate approach.

- It ignores the very localised and uneven distribution of congestion across the city. As a matter of practical policy it is unlikely to be appropriate to impose city-wide solutions to a localised problem.

Furthermore, even accepting the relevance of city-wide averages, aggregate results will give highly biased estimates of the average values of the strongly non-linear effects of congestion.

- By implicitly treating urban travel as a market for vehicle-kilometres, it assumes that travellers' only options are to consume more or fewer vehicle-kilometres. In fact, people have many other options (Hills 1993). In addition to deciding to reduce their travel, they may also change their
  - route;
  - mode;
  - destination;
  - departure time; or
  - vehicle occupancy.

  While these options may indeed be reflected in vehicle-kilometres travelled, the relationship is not simple, and may even be opposite to that expected. Depending on the location of origins and destinations, some travellers might, for instance, respond to a congestion charge by choosing longer routes or more distant destinations (resulting in more vehicle-kilometres travelled). This would imply that, for these travellers, the demand curve was rising.

  The problem is inherent in the assumption that undifferentiated vehicle-kilometres are an adequate measure of urban travel. In fact, travellers do not derive benefit from vehicle-kilometres but from completed trips to their various destinations. For a given trip, the benefit is associated with 'getting from A to B' and, other things being equal, travellers would generally choose the shortest route. This suggests that vehicle-kilometres are more appropriately seen as an intermediate input whose costs are to be minimised than as the commodity in final demand.

  Redefining demand in terms of trips does not resolve the problem for aggregate analysis because trips can be of many different types and incur different costs.

- It is unclear how the cost curves should be determined on a network. They depend, ultimately, on the details of traffic flow and vehicle operations as outlined in chapter 3. Plausible forms for these curves can be obtained for single network links where the concepts of traffic flow and capacity are reasonably clear. These forms were discussed in chapter 3, and appendix I gives the forms for the road types considered in this report. But no useful extensions of the concepts of traffic flow and capacity to the whole network have been found which do not also assume that traffic patterns are fixed.

APPLYING THE THEORY TO TRAFFIC ON INDIVIDUAL LINKS

Urban transport models provide a means of overcoming many (but not all) of these difficulties. As described in chapter 2, such models use trips as the basic unit of analysis and deal explicitly with trip generation and the choice of destination, mode, route and vehicle occupancy. Only the choice of departure time is not explicitly included. The resulting travel demand is assigned to routes on the road and public transport networks on the assumption that travellers choose routes to minimise total travel cost. Model output provides the traffic volumes and speeds *on every link of the road network.*

Urban transport models therefore allow the theory to be applied to each link individually. The commodity in demand is simply passage along a specific link, and the average and marginal cost curves, determined by traffic engineering parameters, relate to the flow of traffic on the link. As far as these quantities are concerned, questions of traffic patterns on the network and the range of choice open to travellers do not arise. But the model's estimate of the traffic flow on the link *takes account of the overall demand for travel, the range of options open to travellers, and traffic flows on all the other links*.

The theory still needs a demand curve. Demand for travel on a particular link depends on the basic demand for travel, on the desired trips, on mode choice, and on the availability of (and congestion on) other links. While the models deal with all these issues, it is not possible to plot out the demand curve as a simple function of traffic on the link. But the models *do* tell us the equilibrium traffic flow on each link. This is the quantity OD in figure 4.1. As the cost curves are known, this traffic flow also determines the points A and B in the figure. By definition, of course, the point B is also on the demand curve.

These results refer to current traffic flows, and are obtained from models calibrated to simulate actual travel behaviour as closely as possible. At the level of traffic assignment and route choice it is generally assumed that travellers try to minimise their expected travel times under the prevailing conditions—that is, they make their decisions on the basis of the average cost curve in figure 4.1.

It is, however, possible to simulate the traffic flows that would result if travellers' decisions also took account of the costs imposed on other travellers. All that is required is that, in the traffic assignment step of the modelling process, the calculation of the costs associated with each link should be based on the *marginal* rather than the *average* cost curve in figure 4.1.

It is simply a matter of altering the computer code in the assignment algorithm to replace the formula for the average user cost with the formula for the marginal user cost. Running the full iterative model with this 'marginal-cost' assignment algorithm causes the increase in cost to feed back to trip generation, destination choice, and choice of route. The results include a new O–D matrix, a new skim file, and new equilibrium traffic flows, which reflect the response of travellers to the full social costs of their travel decisions.

The resulting equilibrium traffic flows are those that would be obtained if travel decisions *did* take account of congestion costs imposed on others. These are the socially optimal traffic flows, represented for each link by the quantity OJ in figure 4.1. As before, the cost curves being known, this traffic flow determines the points E and G in the figure. By definition, the point E is also on the demand curve, and the quantity GE is the optimal congestion charge for the link.

To summarise: urban transport models allow the economic theory of congestion to be applied at the level of individual links (which is where the impact of congestion is experienced) while still taking account of the complexities of people's travel choices and of traffic operations on road networks. The application involves:

- known average and marginal travel cost functions on every link;
- simulating current traffic patterns on the assumption that people take account only of their own expected costs as determined by the average cost curve on each link. This establishes the current quantity of travel OD, and the point B at the intersection of the demand and average cost curves; and
- simulating optimal traffic patterns on the assumption that people take account of the full social costs of their travel decisions as determined by the marginal cost curve on each link. This establishes the optimal quantity of travel OJ, the point E at the intersection of the demand and marginal cost curves, and the point G on the average cost curve.

All the important elements of the economic analysis (the cost of congestion, the charge, the revenue, consumer surplus losses, and the net benefit) can therefore be determined, and summed or averaged as required over the whole city.

Theoretical concerns

Urban transport models determine an equilibrium between the demand for travel, which depends on the zonal distribution of facilities throughout the city, and the costs of travel, which are incurred on individual links. The basis for the use of these models in the economic analysis of congestion, as outlined above, is that the overall costs and benefits of urban travel (and estimates of the city-wide measures of congestion) can be obtained by summing the costs and benefits on each link.

There appears to be no problem about summing costs. The situation with benefits is not so clear.

The modelling process makes no use of any demand curve representing the benefits of travel on individual links. This is because the benefits of travel are associated with whole trips and these are dealt with elsewhere in the model. However, (within the limitations of their assumptions) the models correctly estimate link traffic flows with and without congestion charges[6] and, for the purposes of the economic analysis, it is tempting to identify the appropriate corresponding points on the costs curves with points on the demand curve for travel on the link. Thus B (in figure 4.1) is the intersection of the demand and average cost curves, and E the intersection of the demand and marginal cost curves.

The problem is that demand curves are defined *other things being equal*, and in this case other things, most notably the traffic flows on other links, are not equal. The demand curve passing through E is therefore not the same as the one passing through B, and estimates of consumer surplus based on the assumption that it is are not strictly valid.

The consequence is that, *for any particular link*, the estimate of the net benefit of imposing the optimal congestion charge may be in error. However, the error can be in either direction and, when summed over a large number of links, much of it will cancel out. It is unlikely, therefore, that this problem is a major source of error for sums or averages over the whole city.

6. More precisely, with traffic assignment based on the marginal or the average cost of travel.

It should be emphasised that these errors arise *only* for estimates of economic quantities represented by areas bounded by the segment EB of the demand curve. Estimates of the cost of congestion, the optimal charge and the revenue from the optimal charge are theoretically correct for each individual link (within the limitations of the model).

CONGESTION ESTIMATES BASED ON AGGREGATE ANALYSES

Business Council of Australia

As part of an examination of road pricing prepared for the Business Council of Australia, Meyrick (1994) used data from Commeigne's (1992) study of the costs of congestion in Sydney to estimate the charge, revenue and net benefit from imposing a uniform congestion charge over the whole of the Sydney metropolitan area.

He found:

- an optimal uniform congestion charge of 7.6 cents / veh-km;
- annual revenue from this charge of $1.92 billion; and
- annual net benefit from the imposition of the charge of about $20 million, or about 1 per cent of the charge revenue

The main points of the calculation are summarised in appendix II There are three reasons for the very low estimate of net benefit.

(i) As Meyrick shows elsewhere in his report, if the level of congestion varies over an area, the net benefit to be obtained from a uniform charge is necessarily less than could be obtained from a varying charge optimised for each level of congestion. Moreover, the greater the variability of congestion, the greater the loss in net benefit with a uniform charge.

Congestion varies very greatly over an urban road network. The net benefit from the uniform charge might therefore be substantially reduced from the full benefit available with a varying charge. There may, nevertheless, be considerable interest in a uniform charge because of its greater practical and administrative simplicity. In fact, this is precisely Meyrick's concern. The question that then arises is whether the aggregate analysis provides an unbiased estimate of the net benefit from a uniform charge.

(ii) Congestion effects are highly non-linear in the quantity of travel. The major contribution to the total cost of congestion in a city comes from a relatively small proportion of the network where the marginal cost of travel (as well as the benefits from reducing it) is very high. Attributing the same total cost to a uniform level of congestion, as is implicit in the aggregate analysis, reduces the proportion of travel for which the costs greatly exceed the benefits. As the net benefits from congestion charges consist in the elimination of such travel, estimates of these benefits by an aggregate analysis are biased downwards.

It should be noted that Meyrick derives a result which appears to show that it is legitimate to estimate the net benefit from a uniform charge as if the congestion were uniformly distributed across the network. However, this result relies on assumptions of linearity and low congestion which explicitly exclude the effects we are considering.

(iii) The elasticity of demand for travel is assumed to be -0.2. This is a representative value of the demand elasticity with respect to travellers' financial outlays but, as discussed in the last section of this chapter, needs to be adjusted for use with the full generalised cost of travel.

In Meyrick's analysis, vehicle operating costs are 37 per cent of total travel costs, the rest being time costs. Assuming that the elasticity of -0.2 is with respect to these vehicle operating costs, the adjusted demand elasticity with respect to full generalised costs (including time costs) would be -0.54. If this value were used, the estimate of the charge and the revenue would not be much changed, but net benefit would be 2.5 times greater.

In fact, this need to adjust the demand elasticity interacts with the neglected variability of congestion to provide an additional source of downward bias to the estimate of net benefit. As was noted in (ii), the assumption of a uniform level of congestion downplays the contribution of congestion peaks where net benefits from charging are greatest. But it is precisely in these peaks that time costs are greatest and financial outlays represent the smallest proportion of generalised cost. The required elasticity adjustments for these peaks would therefore be greater than indicated, but are omitted in the aggregate analysis.

The ARRB Travel Cost Model (ATCM)

The Australian Road Research Board (ARRB) has developed a simple spreadsheet model to provide estimates of urban travel costs which can be readily updated with routinely available data (Hepburn and Luk 1994). A recent version, known as model P, includes a treatment of congestion costs and allows congestion charges to be estimated (Luk 1994).

The model recognises five road types (freeways, CBD streets, inner arterials, outer arterials, and local roads) and four traffic periods (the morning peak, the inter-peak period, the evening peak, and off-peak periods). The basic unit of analysis is the aggregate annual traffic flow (in vehicle-kilometres) on one of the road types during a particular period of the day, and the standard aggregate congestion analysis is applied to this unit. The results from each unit can then be combined to provide annual averages and totals for the whole city.

Details of the model are provided in appendix III.

The strength of the ARRB Travel Cost Model is its relative simplicity and modest data requirements. An additional benefit of the simplicity of the model is that it is feasible to systematically test the sensitivity of the results to the main assumptions.

The ARRB model is, of course, subject to the biases of the standard aggregate analysis of congestion—at least so far as the basic unit of analysis (combination of road type and traffic period) is concerned. To the extent that these units are correlated with different levels of congestion, the biases are reduced. However, the assumption that travel within the different units is independent neglects the partial substitutability of travel during different periods and, more importantly, the complementarity of travel on different road types. Few trips are made on only one road type.

In 1995 the BTCE commissioned ARRB to provide a consistant set of congestion analyses covering the five road types and four traffic periods for Melbourne, Sydney, Brisbane, Adelaide and Perth.

Table 5.1 shows the peak period congestion charges estimated by model P for each road type in the cities covered by the work commissioned for the BTCE. Table 5.2 shows the principal congestion measures for the same cities. Further results, including examination of the sensitivity of the congestion measures to demand elasticity, are provided in appendix III.

TABLE 5.1 PEAK PERIOD CONGESTION TOLLS IN AUSTRALIAN CITIES

(cents per vehicle-kilometre)

	Melbourne	*Sydney*	*Brisbane*	*Adelaide*	*Perth*
Freeways	14	13	14	0	14
CBD streets	57	62	40	40	40
Arterial (inner)	20	21	16	16	16
Arterial (outer)	7	7	5	5	5

Source ARRB Contract Report No. CR TE 95/012.

TABLE 5.2 ECONOMIC MEASURES OF CONGESTION IN AUSTRALIAN CITIES[a]

($million per year)

	Melbourne	*Sydney*	*Brisbane*	*Adelaide*	*Perth*
Cost of congestion[b]	809	926	176	111	133
Congestion charge measures[c]					
Toll revenue	1 608	1 816	384	230	275
Road user losses[d]	1 141	1 295	278	164	191
Net benefit[e]	466	521	107	67	83

a. Summed over all road types (except local) and all traffic periods (except off-peak).

b. The value of time lost due to congestion.

c. As compared with the current congested situation with no charge.

d. Losses to all road users, the tolled and the tolled–off.

e. Net benefit = revenue–road user losses.

Source ARRB Contract Report No. CR TE 95/012.

UNIFORM TOLL ESTIMATES—THE EFFECT OF AGGREGATION

The difference between an aggregate and a link-by-link analysis can be illustrated by examining the aggregated results from a series of TRANSTEP runs which simulate the impact of varying levels of a uniform charge.

The starting point is the calibrated TRANSTEP model of the weekday morning peak hour in Melbourne which is discussed in chapter 6 and used to obtain the congestion and optimal charge estimates presented in chapter 7. The output from this model is intended to represent as closely as possible the ‘current’ patterns and quantities of travel—with no charge. Recall that the results provide traffic volumes (and, hence, all the congestion measures) on every link of the network. These can be aggregated to provide whatever averages or totals are required.

The model can also be run to simulate travel subject to a uniform charge per vehicle-kilometre. Table 5.3 shows the results from a series of runs with different levels of a uniform charge and, for comparison, the corresponding results for the optimal regime of varying charges, for which the average charge (revenue divided by total travel) is 17.2 cents per veh-km.

The average cost of travel is the total cost divided by the total travel. The total cost is made up of the costs of vehicle maintenance, fuel (which is a function of both speed and volume–capacity ratio), travel time, plus the amount paid in charges.

Figure 5.1 shows the average cost of travel *including and excluding the charge* (from table 5.3) plotted against total travel. These results, which

TABLE 5.3 RESPONSE TO A UNIFORM CONGESTION TOLL – COMPARISON WITH REGIME OF OPTIMAL TOLLS

	Average cost of travel				
Toll	*including charge*	*excluding charge*	*Total travel*	*Net benefit*	*Toll revenue*
($/veh-km)	*($/veh-km)*	*($/veh-km)*	*(veh-km)*	*($)*	*($)*
Uniform charge					
0.00	0.451	0.451	6 888 287	0	0
0.05	0.436	0.486	6 327 578	135 548	316 386
0.09	0.425	0.515	5 900 286	203 690	531 030
0.13	0.413	0.543	5 505 586	251 577	715 717
0.15	0.409	0.559	5 316 005	254 650	797 395
0.17	0.405	0.575	5 137 046	254 283	873 312
0.22	0.398	0.618	4 705 208	213 358	1 035 148
0.26	0.390	0.650	4 397 678	172 804	1 143 420
Optimal charge regime					
0.17[a]	0.354	0.526	5 782 437	510 300	992 536
Aggregate analysis					
0.13	0.414	0.543	5 501 207	141 765	712 959

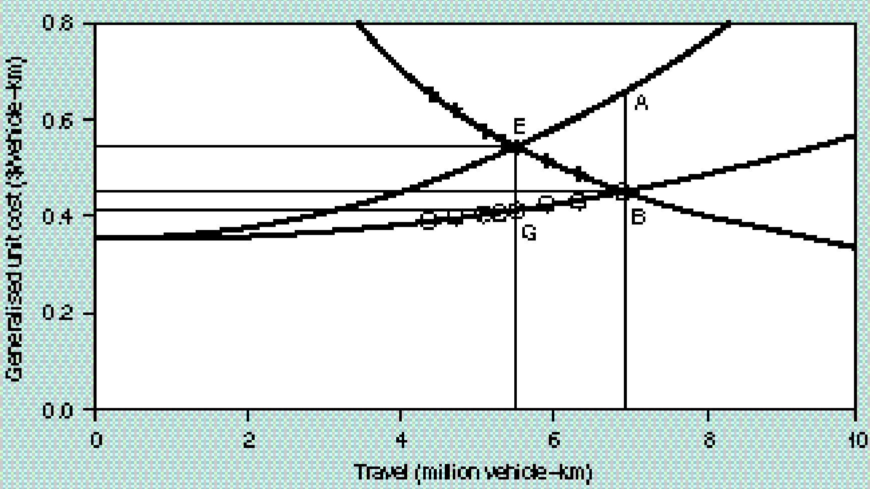
a. Average charge – revenue divided by total travel.

Source BTCE modelling results for the Melbourne AM peak hour.

are separated by the value of the charge which leads to the corresponding quantity of travel, clearly correspond to the demand and average cost curves of the standard aggregate congestion analysis. The figure also shows functions fitted to the two sets of plotted points.[7] Given the function for the average cost curve, the marginal cost curve follows, and this is also shown.

The results of an aggregate analysis based on figure 5.1 are shown in the last line of table 5.3. The marginal cost and demand curves intersect at the point where the total travel equals 5 501 207 veh-km and the separation between marginal and average cost curves is 13.0 cents per veh-km. This would be the estimate of the level of the optimal uniform charge. The product of this charge and total travel gives the charge revenue $712 959. The net benefit is given by the area ABE, which turns out to be equal to $141 765.

FIGURE 5.1 AGGREGATE CONGESTION ANALYSIS USING TRANSTEP MODELLING RESULTS.



Note Points represent average cost of travel including (+) and excluding (O) charge from Table 5.3. Curves fitted to data points. See text.

Source BTCE modelling results for the Melbourne AM peak hour.

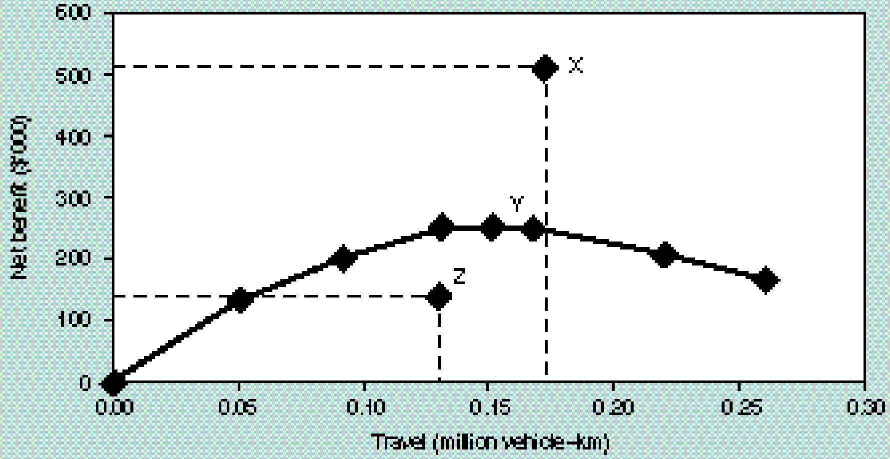
BTCE

7. Using P (in $) to represent generalised unit cost and Q (in millions of veh-km) to represent quantity of travel, the constant elasticity function $Q = 2.60\ P^{-1.23}$ was fitted to the points on the demand (average cost with charge) curve. The function $P = 0.353\ [1 + 0.156\ (Q/5.26)^{2.14}]$ was fitted to the points on the average cost (without charge) curve. The corresponding marginal cost curve is given by $P = 0.353\ [1 + 0.489\ (Q/5.26)^{2.14}]$.

Figure 5.2 shows the net benefit from applying different levels of a uniform charge (results from table 5.3). The maximum value occurs for a uniform charge of about 16 cents per veh-km (Y). The aggregate analysis based on figure 5.1 obtained an optimal uniform charge of 13 cents per veh-km, but substantially underestimates the net benefit (Z). For comparison, the net benefit for an optimal regime of varying charges (which has an average charge of 17.2 cents per veh-km) is plotted at X.

Figure 5.1 represents an aggregate analysis of the (disaggregate) modelling results. The demand and cost curves in figure 5.1 are derived by aggregating link-by-link modelling results. They are not subject to errors of measurement, sampling errors, or inconsistencies. The demand curve correctly represents the model's response to a uniform congestion charge in terms of people's demand for vehicle-kilometres of travel. The average cost curve correctly describes the cost per vehicle-kilometre of the different quantities of travel, taking into account the complexities of travel behaviour and network operations. And yet the net benefit from the optimum level of a uniform congestion charge is still substantially underestimated by the aggregate analysis.

FIGURE 5.2 VARIATION OF NET BENEFIT FOR DIFFERENT VALUES OF A UNIFORM CONGESTION CHARGE



Note The optimal level of the uniform charge (which maximises net benefit) is about 16 cents per veh-km (Y). Aggregate analysis estimates optimal uniform charge as 13 cents per veh-km and much lower net benefit (Z). For comparison: optimal regime of varying charges has average charge of 17.2 cents per veh-km and higher net benefit (X).

Source BTCE modelling results for the Melbourne AM peak hour (table 5.3).

BTCE

In actuality, an aggregate analysis could not do as well as this. The error-free demand and cost curves in figure 5.1 were obtained by aggregation of the link-by-link results from the TRANSTEP model, and would not have been otherwise available.

ELASTICITY OF DEMAND WITH RESPECT TO GENERALISED COST

The elasticity of the demand curve in figure 5.1 is -1.23. Estimates of the elasticity of travel demand vary considerably, but this is substantially above (more elastic than) the generally accepted values. Luk and Hepburn (1993), in a review of Australian travel demand elasticities, noted values of -0.10 in the short run and -0.26 in the long run for the elasticity of traffic levels *with respect to the price of fuel*.

The cost of fuel, however, is only a small part of the overall cost of travel. A given fractional variation in its price therefore represents a much smaller fractional variation in the overall cost of travel, and the elasticity of demand for travel *with respect to its full cost* is correspondingly greater than that *with respect to the price of fuel*.

Demand elasticity (ε) is defined as the ratio of the fractional change in quantity demanded (Q) to the fractional change in the price (P) thought to be responsible for the change in quantity, other things remaining equal.

$$\varepsilon = \frac{\Delta Q}{Q} \Big/ \frac{\Delta P}{P} = \frac{\Delta Q}{\Delta P} \frac{P}{Q}$$

If P is the full price of the commodity in question, then ε is the normal own price demand elasticity. But the elasticity can be defined with respect to a price which is only part of the whole, as was done with fuel price above. (Note that $\Delta P_{\text{part}} = \Delta P$ because no other component of price changes.)

$$\varepsilon_{part} = \frac{\Delta Q}{\Delta P} \frac{P_{part}}{Q}$$

The full own price elasticity can then be obtained by adjusting the part price elasticity.

$$\varepsilon = \varepsilon_{part} \frac{P}{P_{part}}$$

The concept of generalised cost, which is central to the economic analysis of congestion, includes the conventional financial costs such as maintenance and fuel costs but also includes, and is usually dominated by, the cost of travel time. The fuel costs (with respect to which the elasticity is known) are therefore only a part of the full generalised cost. In the case of Meyrick's analysis, for instance, it was noted that non-time costs were only 37 per cent of full generalised costs, and it was therefore suggested that the financially based estimate of demand elasticity be multiplied by 2.7 (= 1/0.37). In fact, greater adjustment might well have been appropriate to take account of the elimination of congestion peaks in the aggregate analysis.

In fact, choice of elasticity is not a problem when the analysis is applied to individual links, because elasticity is not an input to the process. The models can, of course, be used to determine various elasticities implied by the embodied behavioural assumptions. The demand curve with elasticity -1.23 in figure 5.1 is one example.

A set of runs of the Melbourne model was performed in which the price of fuel was varied systematically. These produced results that were consistent with an elasticity of demand for travel with respect to the price of fuel of -0.1. Since Luk and Hepburn report an empirical estimate of this elasticity as -0.1 in the short term, and -0.26 in the long term, it appears that the elasticity assumed in these models is, if anything, rather low. It must be noted, then, that in chapter 7 the estimates of road user charges are upper bounds, and those of net benefits are lower bounds.

CHAPTER 6 CALIBRATION

The results in this report depend on a modelling process, which in turn depends upon data. By far the greater part of this data is collected from direct and explicit sources. But some of the data, chiefly crucial parameters in various behavioural models, are inferred from aggregate data by an iterative process called 'calibration'.

Calibration can seem an arcane procedure. Calibration is also often seen as completely ad hoc, a mere large-scale 'fudging' of results. It is therefore useful to explain the process that underpins the results in this report.

The need for calibration

As explained in chapter 2, urban traffic models produce their estimates of the overall level and pattern of traffic in a city by dividing the city into hundreds of 'zones', and then simulating the travel choices of small groups on a local basis. That is, they simulate the decision to travel ('trip generation') zone by zone. They simulate the choice of destination ('trip distribution') zone by zone. And they simulate the choice of public or private transport ('mode split') and the choice of route ('trip assignment') by origin–destination pair.

Obviously, this process requires that the data needed for modelling be available on a disaggregated, zone-by-zone basis. For the most part, this is not a problem.

- The land use data discussed in chapter 2 are readily available in suitably disaggregated form; for example, from the census.
- The costs of travel by various modes and routes between all origin–destination pairs are estimated within the 'assignment' step, and stored in 'skim matrix' files for use in the 'distribution' and 'mode split' steps.

- Data on the road network and public transport routes are not readily available, but can be obtained and collated by patient research, and are available from transport planners.

What is missing is the quantitative link between these populations, opportunities and costs on one hand, and the amount and pattern of actual travel, on the other. One may be confident that the basic demand for travel from a zone is proportional to its population. But how many trips should there be per person? High generalised costs of travel between two zones should result in few trips taking place between them. But how strong is this effect? The mode split ratio between public and private transport should depend upon the relative generalised costs of travel by the modes, and the length of the trip. But how strongly does the effect depend on these things?

Some data that are relevant to these questions are available—for example, the total population of the city and the total number of trips, which imply an average trip generation rate for the city overall. The average lengths and average durations of trips actually undertaken by public and private transport are available from household surveys of expenditure and travel behaviour. Generalised costs are easily calculated from these. The overall mode split ratio for commuter trips in the city, and for trips to the central business district, are available from the same and other similar sources. There are two problems.

(i) The data are aggregates over the whole city or large parts of it, but the model performs simulations in much greater detail.

(ii) The data tells us how people behave when faced with the actual opportunities and costs presented by the city, including all its peculiarities. The model must simulate how people representative of the population of the city would behave in more general circumstances.

In short, the available data correspond to statistics that summarise modelling *outputs*. It is not possible to calculate input parameters from these data in any straightforward way.

The only way to proceed is to try plausible values of the input parameters, and to compare the aggregate results of the simulation with the corresponding aggregate real-world data. If there is a discrepancy, the person calibrating the model makes an educated guess about what change in an input parameter will reduce it, adjusts the parameter, and repeats the simulation. The process can be repeated iteratively until the model produces realistic results.

This procedure is called calibration. Essentially, it is an empirical process of inferring the values of (input) behavioural parameters from emergent statistics (corresponding to output) by trial and error.

Calibration of TRANSTEP models

The BTCE's work in this report required that calibration techniques be used in three TRANSTEP sub-models: the 'distrim' model (which performs both trip generation and distribution), the mode split model and the assignment model.

Calibration of road type characteristics

The TRANSTEP assignment model does not contain any parameters that are normally calibrated. However, the present study was performed on a strategic scale with road networks that were somewhat skeletal or schematic. Because of this, it was found to be impossible to assign the full observed traffic on the somewhat reduced network without producing excessive congestion. Excessive congestion resulted in models that seriously understated the average speed of traffic in the cities.

To overcome this problem, the capacities of different road types were adjusted as follows. The capacities of freeways were not altered from the capacities assumed by traffic engineers. (All freeways were fully represented in the networks used.) The capacities of CBD streets were increased by a variable factor (CBD streets were under-represented in the networks used in this study by a considerable extent). The capacities of all other road types were adjusted to preserve the original relationships between the capacities (the higher the road type capacity, the more completely roads of that type were represented in the networks).

A value was chosen for the scale factor applied to CBD street capacities, and this was used to build a model of Melbourne. This model was calibrated to produce the correct trip numbers, trip costs and mode split characteristics. Then the average traffic speed in the model was compared to the actual average traffic speed in Melbourne. Calibration techniques, with the scale factor adjusted to bring the predicted and observed traffic speeds for Melbourne into agreement, were used.

This calibration involved one free (calibration) parameter: the amount by which the lowest-capacity road types were scaled up. It involved a single target: the average speed of traffic. A satisfactory calibration required doubling the capacity of CBD streets (and interpolating other street type capacities).

The exercise was performed only once, using a model of Melbourne, and the road type capacities thus determined were used for all six cities. Results were satisfactory.

Calibration parameters in trip generation and distribution

The TRANSTEP distrim model has four parameters that must have their values determined by calibration. Essentially, these define the position and shape of a general demand curve for travel. This is scaled by the population of each origin zone, and convolved with the travel opportunities from that zone to produce a demand for travel from that zone.

The shape of the aggregate demand curve for travel had to match two criteria:

(i) that the demand curve for aggregate travel should have a constant elasticity of total vehicle-kilometres with respect to *generalised cost of travel*;

(ii) that the elasticity of total vehicle-kilometres with respect to the price of fuel should be consistent with findings reported by other researchers.

The shape parameters of the ‘distrim’ curve were calibrated to satisfy these criteria.

The result was an implied elasticity of travel demand with respect to fuel price of -0.1, which is at the conservative end of the range of empirical estimates. Luk and Hepburn (1993) noted values of -0.10 in the short run and -0.26 in the long run.

The elasticity of travel demand with respect to the total generalised cost of travel emerged as -1.2. When increases in the generalised cost of travel were modelled by imposing uniform per-kilometre tolls on all automotive travel, fit to the constant elasticity demand curve for aggregate travel was extremely good.

Three ‘distrim’ parameters were fixed by this calibration of the shape of the demand curve. One free parameter remained in the distrim model, which controlled the position of the demand curve.

Calibration parameters in mode split

The TRANSTEP mode split model allows the use of seven parameters that must be determined by calibration. These are the coefficients on the explanatory variables in a logit model of mode choice. Four of the seven were used in this study, the others were left null. The explanatory variables corresponding to the parameters used were: the length of the trip, the ratio of the (generalised) costs of travel by the different modes, the difference between the costs of travel by the different modes, and the employment density of the destination (a proxy for parking difficulties).

Calibration targets

Available data provided six independent targets that model output had to match: total number of trips, average car trip length, average car trip duration, average generalised cost of a public transport trip, overall mode split ratio, mode split ratio for trips to the central business district.

However, only five input parameters were available for calibration (one in the 'distrim' model and four in the mode split model). The calibration process therefore had fewer control parameters than target statistics. Such a situation is termed 'an over-determined system', and cannot always be solved.

Despite the over-determination of the system, calibration of TRANSTEP models of the cities in this study did not prove to be impossible. The calibration allowed very good results for all cities except Sydney, and good results for Sydney. This provides some reassurance as to the basic soundness of the TRANSTEP modelling process and the results obtained in this study.

Sensitivity analysis

Because the system is over-determined, there were no free variables on which a sensitivity analysis could be performed. Essentially, all the parameter values in these models were dictated by observation, the state of the available network data, and the choice of the functional form of the speed–flow curve.

CHAPTER 7 CONGESTION IN AUSTRALIAN CITIES

TRANSTEP presents its results as traffic volumes and traffic speeds on each link in its road network. This is not very illuminating for a strategic view of the congestion situation over an entire city. Therefore the results of the BTCE modelling project are presented in this chapter in two different forms which summarise the link-by-link results on different bases.

To indicate the scale of the congestion problem in each city, and to permit a rough comparison between cities, whole-city statistics are presented in tabular form.

To provide a strategic view of each city, charts are used to show the spatial distribution of road capacity, travel (vehicle-kilometres travelled), travel time, delay, and the level of optimal charges. These charts are oblique perspective views of 3-dimensional graphs, and are constructed in the following manner.

A grid of square cells, 3-km on each side, is superimposed on the area included in the model of the relevant city. A corresponding grid is constructed in the graph. If there are no roads in a 3-km cell, the corresponding square in the graph is coloured grey. Otherwise, a white column is erected over the square in the graph, with its height proportional to the amount of whatever is being graphed that is or occurs in the corresponding 3-km cell. The resulting 3-D graphs look like clusters of closely spaced tower buildings. Then the oblique perspective view of the graph is drawn. The graphs of Adelaide, Canberra, Melbourne and Perth are drawn as they would appear from the south-west. To facilitate recognition of the coastline, the graphs of Sydney and Brisbane are drawn as they would appear from the south-east.

TABLE 7.1 CONGESTION AND OPTIMAL ROAD-USER CHARGES IN AUSTRALIAN CITIES—MORNING PEAK HOUR

	Units	*Adelaide*	*Brisbane*	*Canberra*	*Melbourne*	*Perth*	*Sydney*	*Total*
Current 'cost of congestion'	$	157606	513639	65451	919698	268432	1074692	2999519
Average level of optimal charges	$ per km	0.03	0.08	0.06	0.17	0.04	0.13	0.10
Maximum level of optimal charges	$ per km	0.08	0.31	0.40	1.26	0.28	0.75	1.26
Revenue from optimal charges	$	71172	249913	72669	992536	109467	555424	2051181
Motorist net losses under optimal charges	$	26673	94860	42142	482235	49317	255723	950950
Net benefit of optimal charges (the cost of doing nothing about congestion)	$	44499	155053	30527	510301	60150	299701	1100231

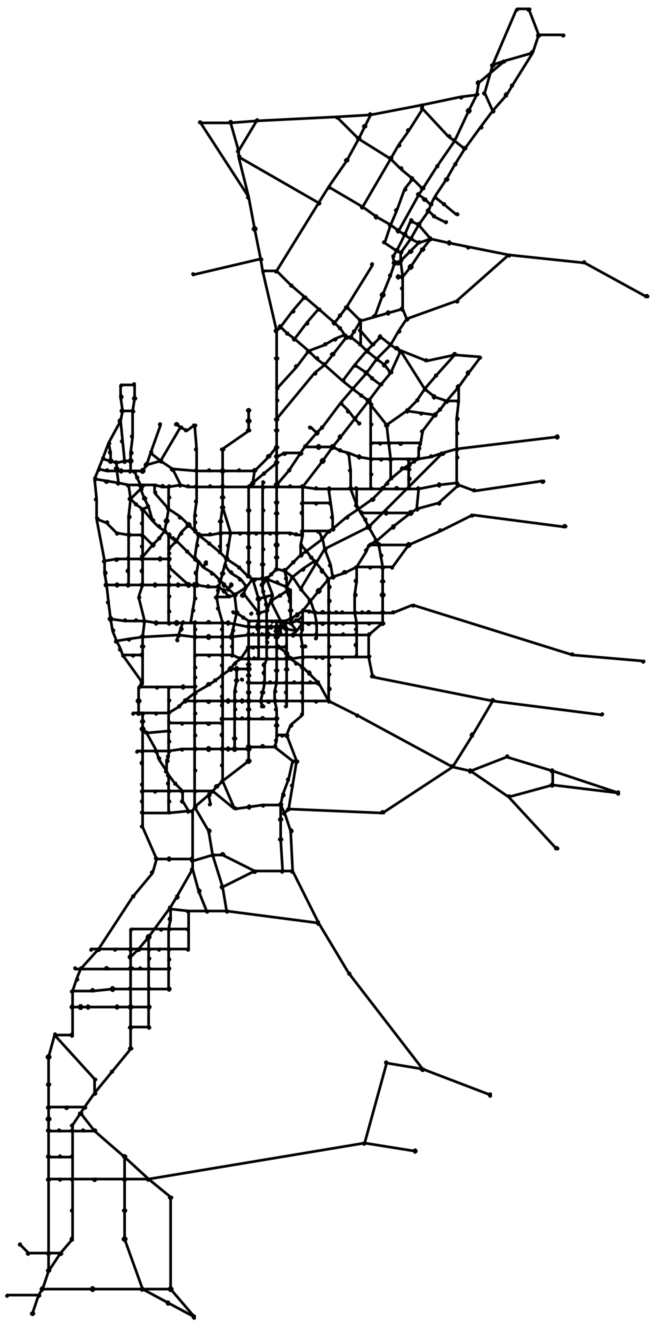
Source BTCE modelling results.

TABLE 7.2 EFFECTS OF THE OPTIMAL ROAD-USER CHARGES ON TRAFFIC CONDITIONS IN AUSTRALIAN CITIES—MORNING PEAK HOUR

	Units	*Adelaide*	*Brisbane*	*Canberra*	*Melbourne*	*Perth*	*Sydney*	*Total*
Travel (current)	Vehicle–km	2836999	3999234	1244191	6888287	3469106	5887569	24325386
Travel (with charges)	Vehicle–km	2450574	3322726	1156741	5782473	2598624	4306363	19617501
Change	per cent	-14	-17	-7	-16	-25	-27	-19
Travel time (current)	Vehicle–hours	67838	117749	27636	193952	86303	211650	705128
Travel time (with charges)	Vehicle–hours	49544	72587	22824	120492	53093	109548	428088
Change	per cent	-27	-38	-17	-38	-38	-48	-39
Delay (current)	Vehicle–hours	10681	35577	4396	63055	18252	75747	207708
Delay (with charges)	Vehicle–hours	1821	5012	1071	11584	2230	10824	32542
Change	per cent	-83	-86	-76	-82	-88	-86	-84
Fuel consumption (current)	Litres	251511	383474	107090	671035	312935	616951	2341996
Fuel consumption (with charges)	Litres	204675	273119	94884	498482	214317	372942	1658419
Change	per cent	-18	-29	-11	-26	-32	-40	-29
Network average speed (current)	km–hour	42	34	45	36	40	28	34
Network average speed (with charges)	km–hour	49	46	51	48	49	39	46
Change	per cent	18	35	13	35	22	41	33

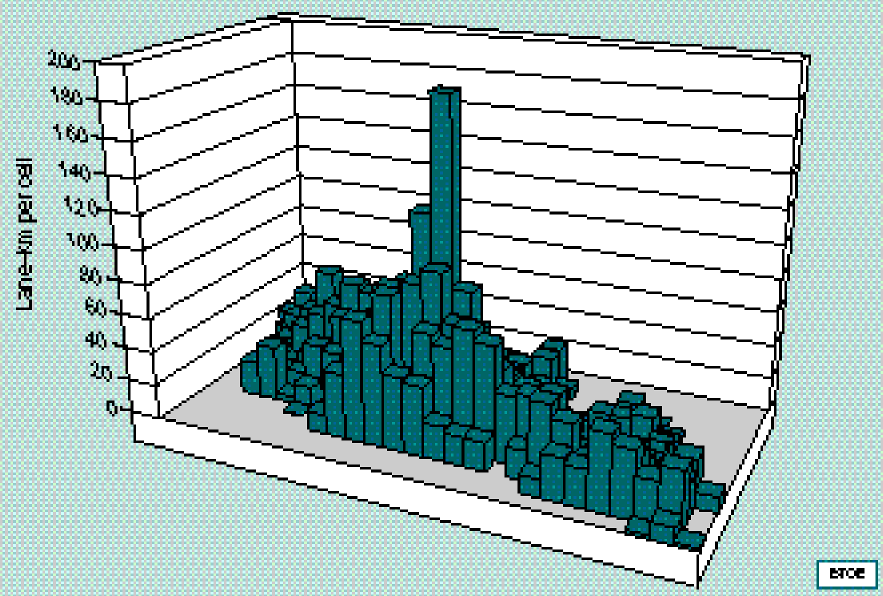
Source BTCE modelling results.

FIGURE 7.1 ROAD NETWORK USED IN THE ADELAIDE MODEL



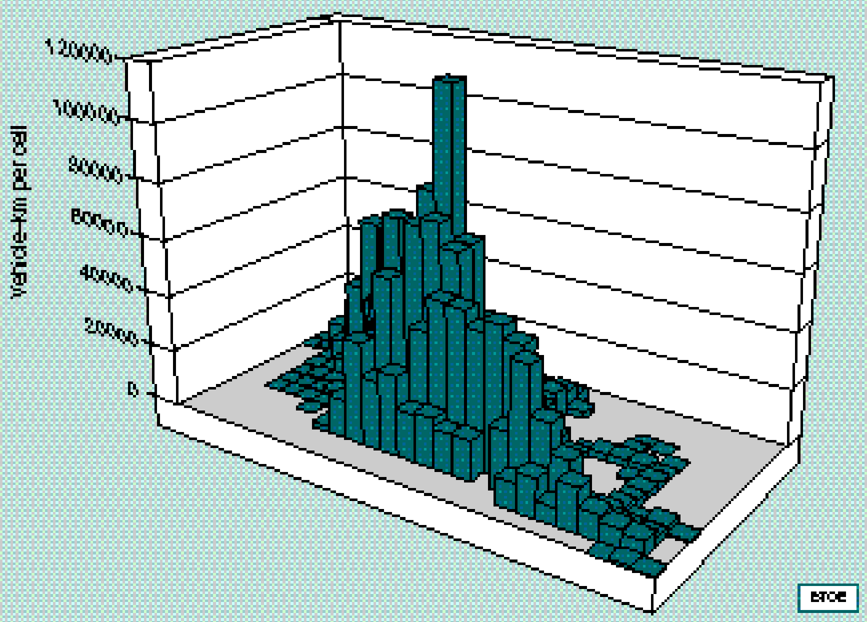
Source BTCE graphical representation of data from RJ Nairn & Partners (pers. comm.)

FIGURE 7.2 ROAD CAPACITY IN THE ADELAIDE MODEL



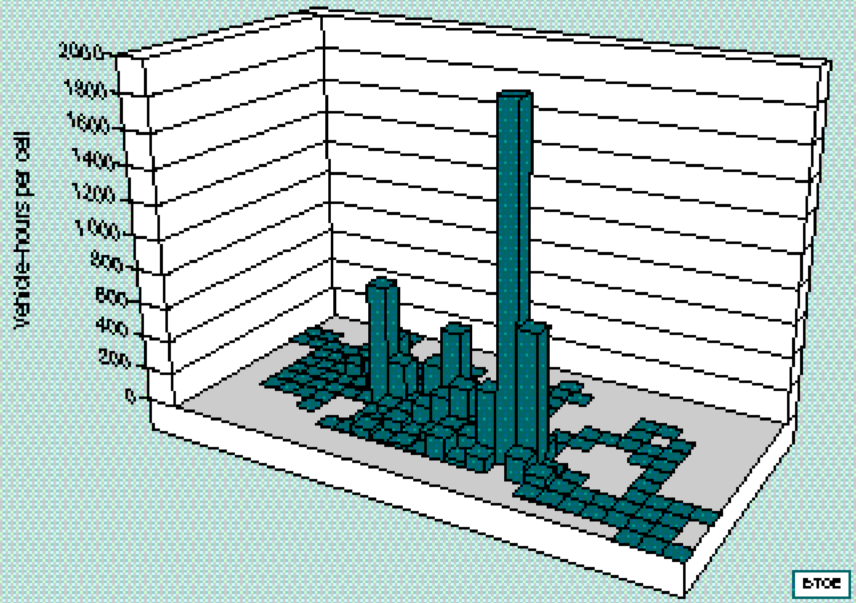
Source BTCE graphical representation of data from RJ Nairn & Partners (pers. comm.).

FIGURE 7.3 TRAVEL IN THE ADELAIDE AM PEAK HOUR



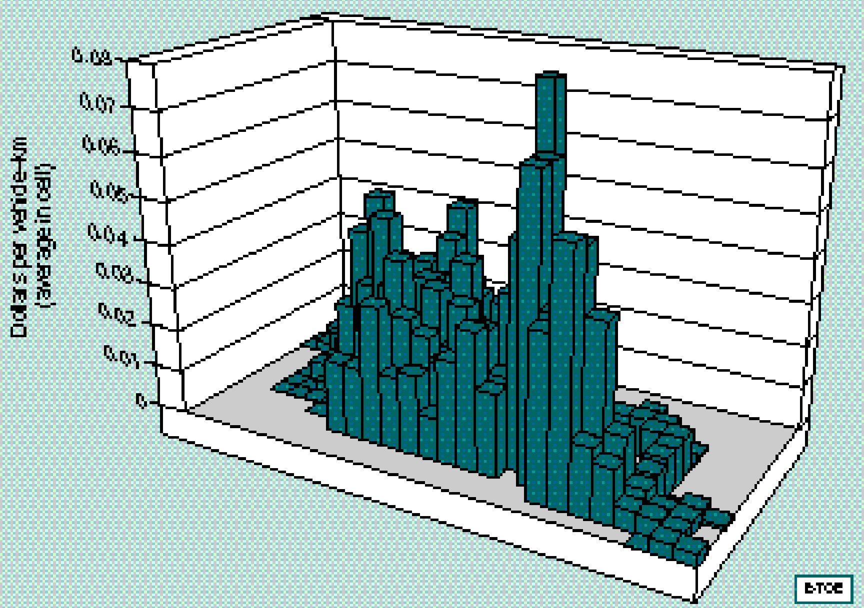
Source BTCE modelling results.

FIGURE 7.4 DELAY IN THE ADELAIDE AM PEAK HOUR



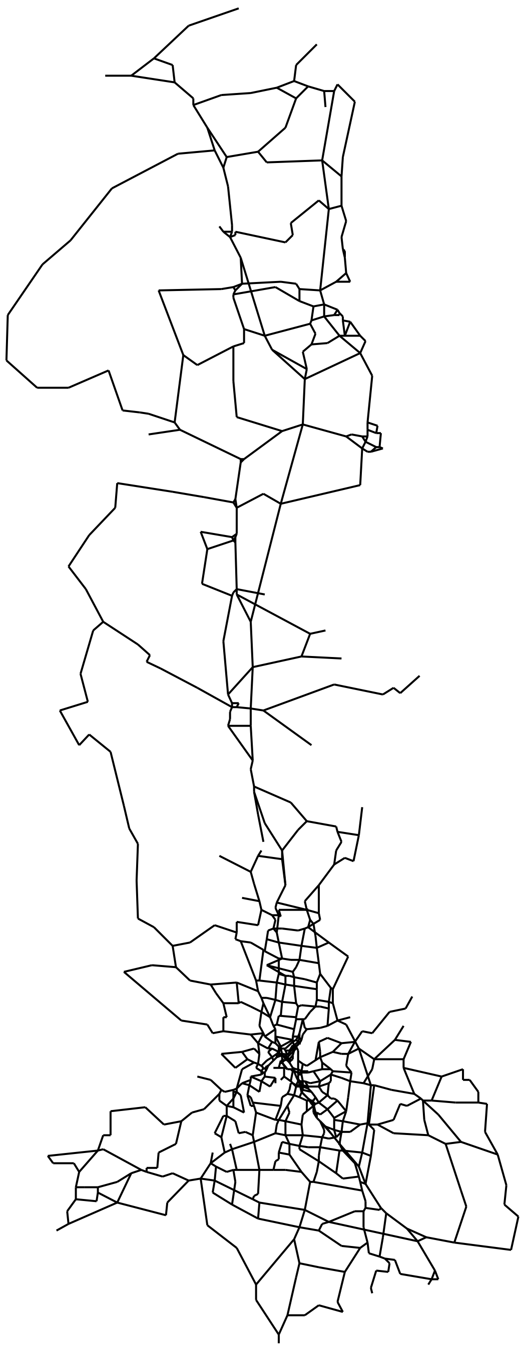
Source BTCE modelling results.

FIGURE 7.5 OPTIMAL ROAD USER CHARGES IN THE ADELAIDE AM PEAK HOUR



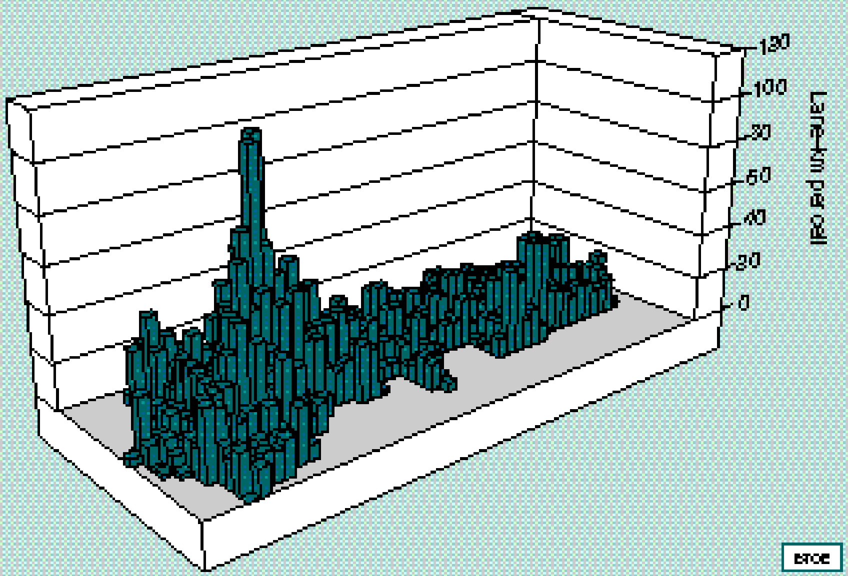
Source BTCE modelling results.

FIGURE 7.6 ROAD NETWORK USED IN THE BRISBANE MODEL



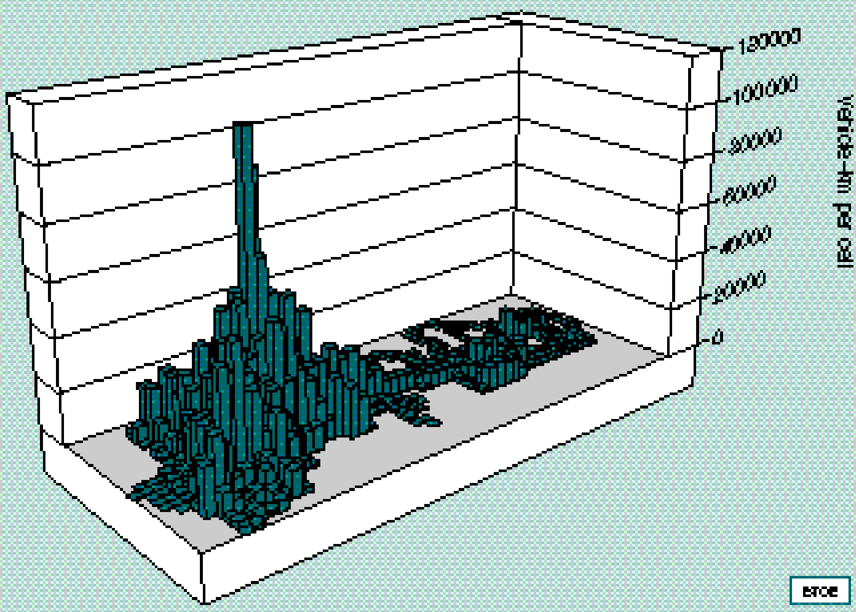
Source BTCE graphical representation of data from RJ Nairn & Partners (pers. comm.)

FIGURE 7.7 ROAD CAPACITY IN THE BRISBANE MODEL



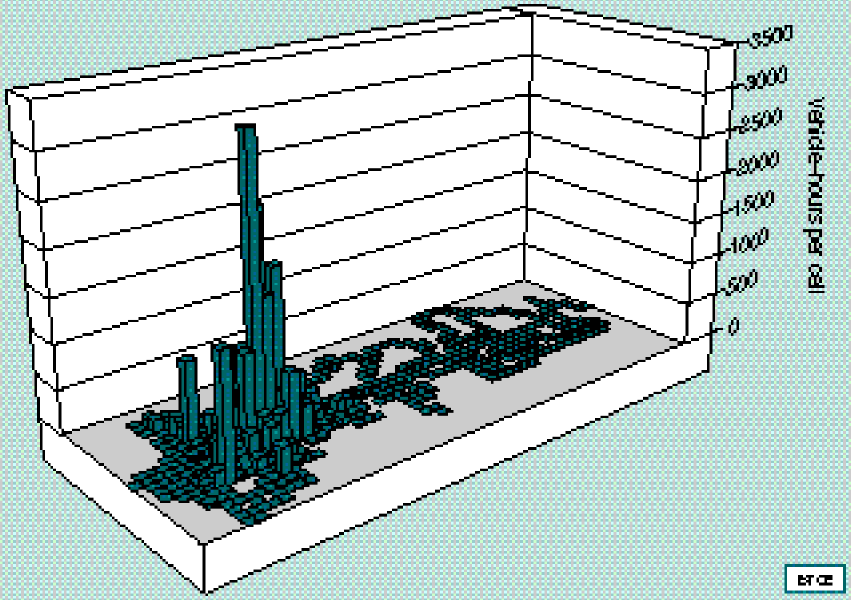
Source BTCE graphical representation of data from RJ Nairn & Partners (pers. comm.).

FIGURE 7.8 TRAVEL IN THE BRISBANE AM PEAK HOUR



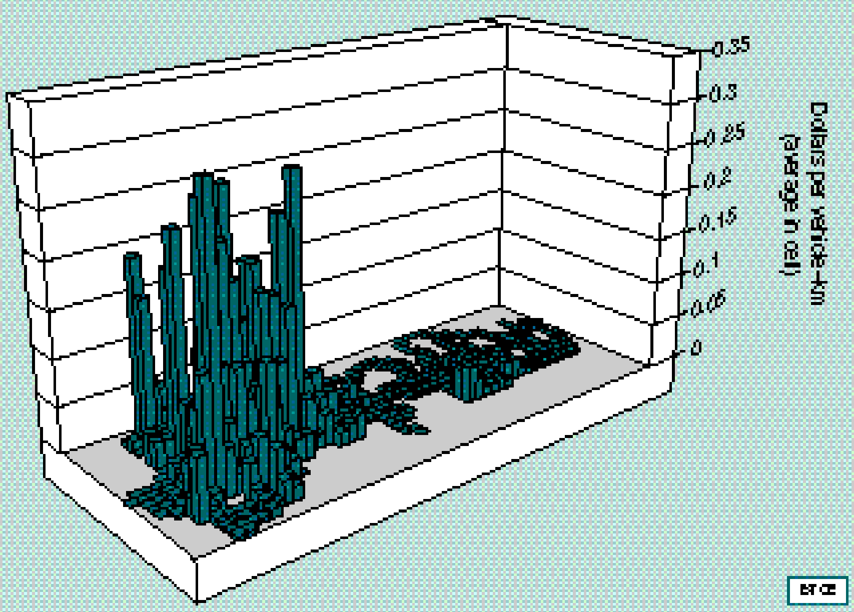
Source BTCE modelling results.

FIGURE 7.9 DELAY IN THE BRISBANE AM PEAK HOUR



Source BTCE modelling results.

FIGURE 7.10 OPTIMAL ROAD USER CHARGES IN THE BRISBANE AM PEAK HOUR

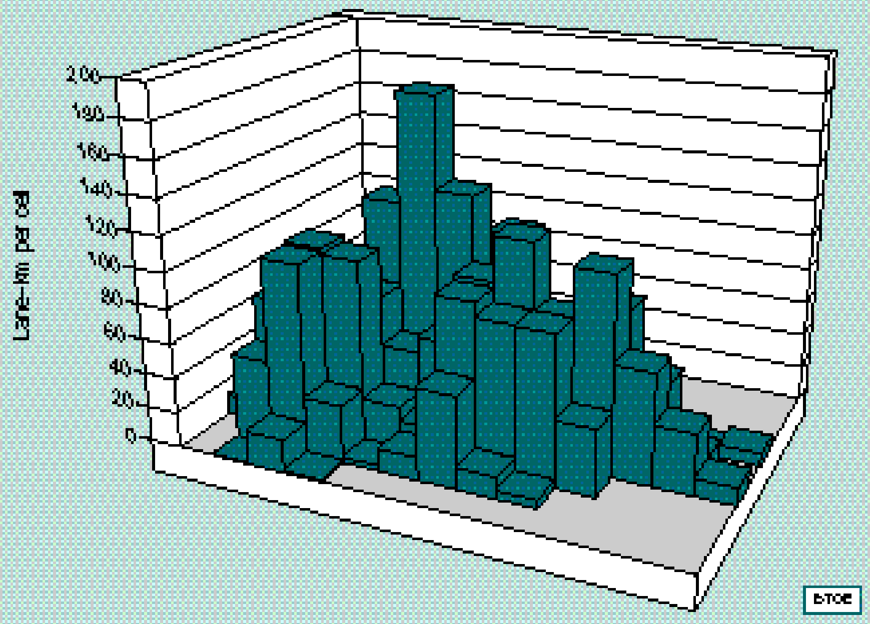


Source BTCE modelling results.

FIGURE 7.11 ROAD NETWORK USED IN THE CANBERRA MODEL

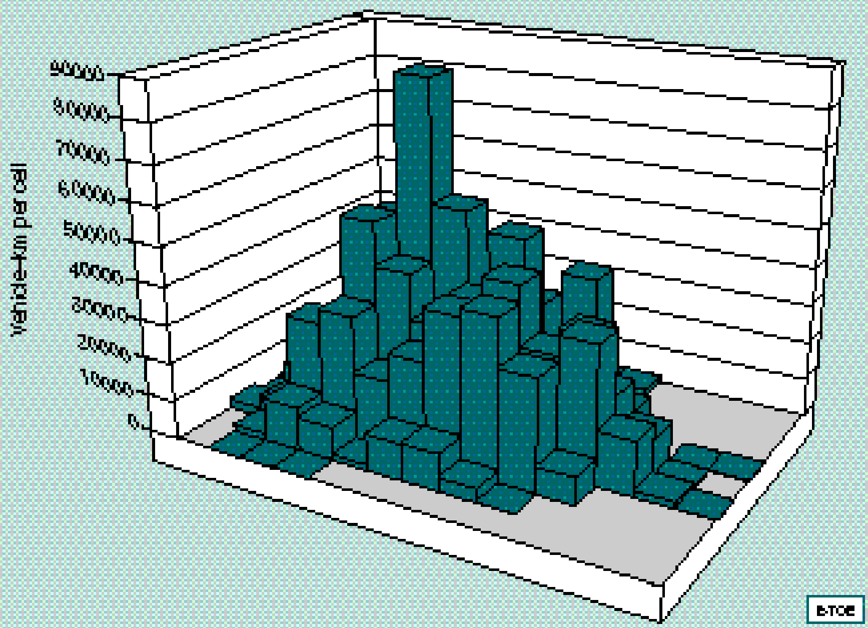
Source BTCE graphical representation of data from RJ Nairn & Partners (pers. comm.)

FIGURE 7.12 ROAD CAPACITY IN THE CANBERRA MODEL



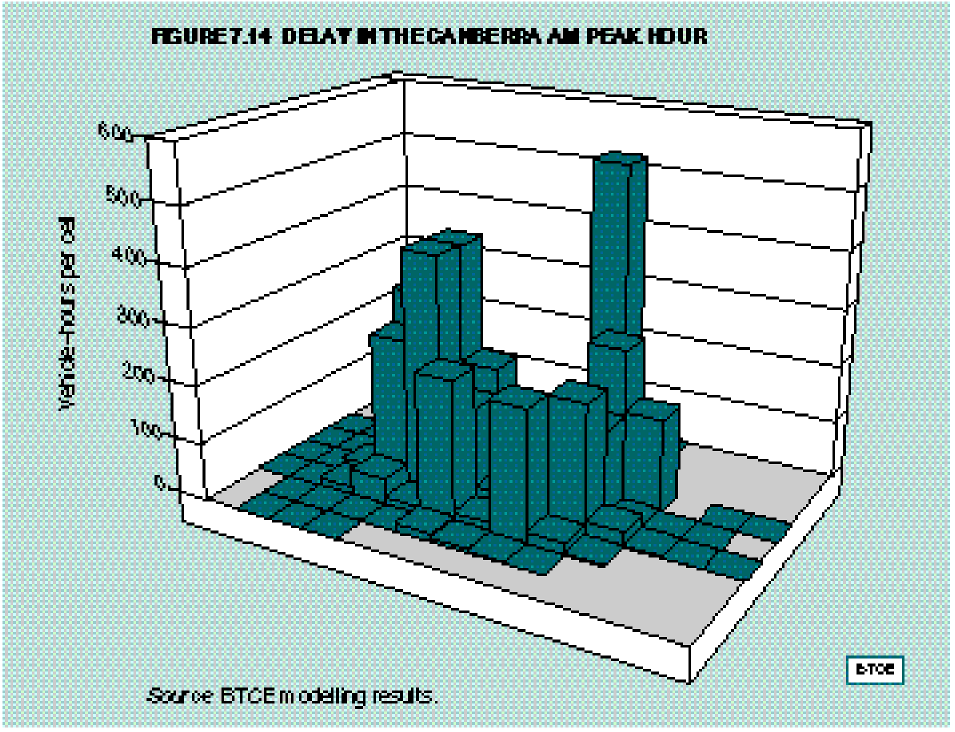
Source BTCE graphical representation of data from RJ Nairn & Partners (pers. comm.).

FIGURE 7.13 TRAVEL IN THE CANBERRA AM PEAK HOUR



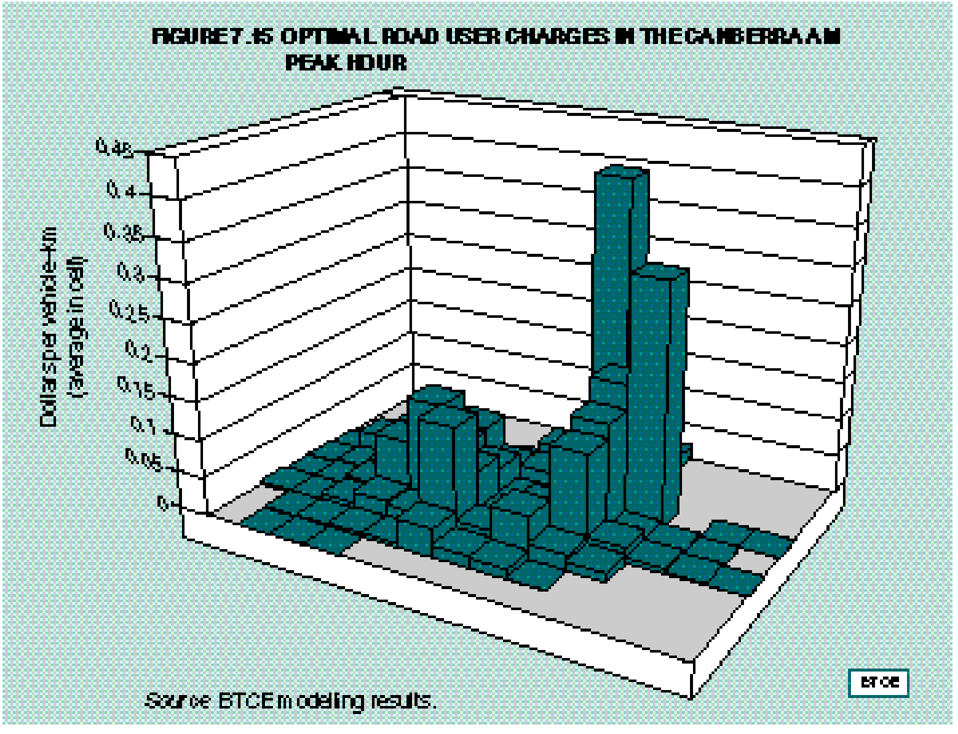
Source BTCE modelling results.

FIGURE 7.14 DELAY IN THE CANBERRA AM PEAK HOUR



Source BTCE modelling results.

FIGURE 7.15 OPTIMAL ROAD USER CHARGES IN THE CANBERRA AM PEAK HOUR

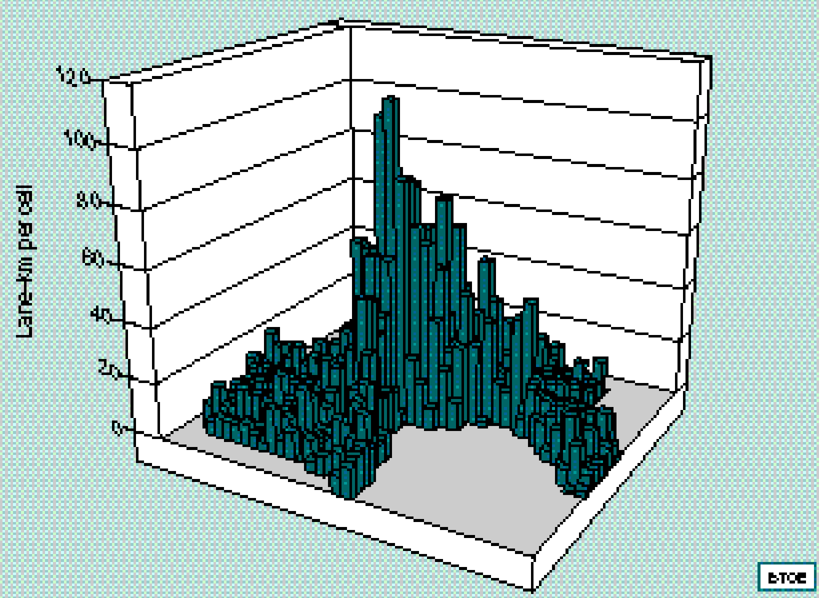


Source BTCE modelling results.

FIGURE 7.16 ROAD NETWORK USED IN THE MELBOURNE MODEL

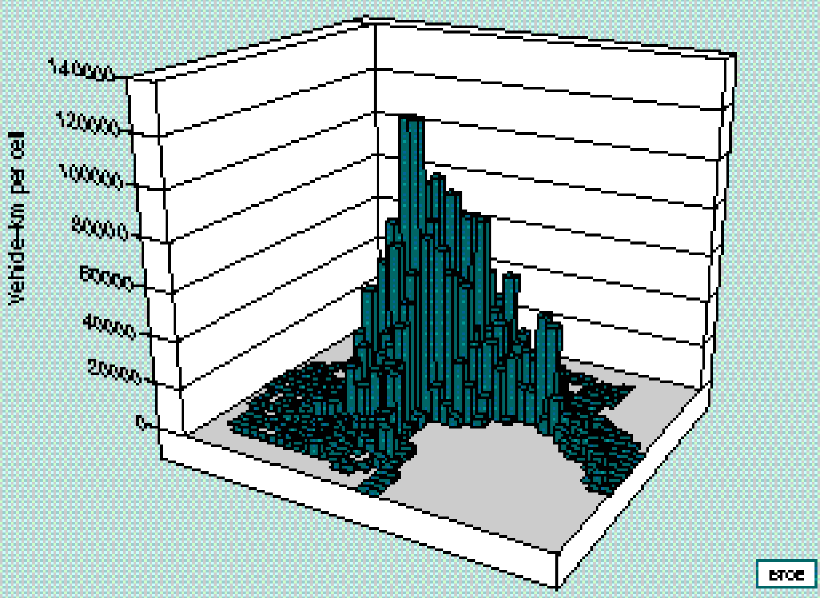
Source BTCE graphical representation of data from RJ Nairn & Partners (pers. comm.)

FIGURE 7.17 ROAD CAPACITY IN THE MELBOURNE MODEL



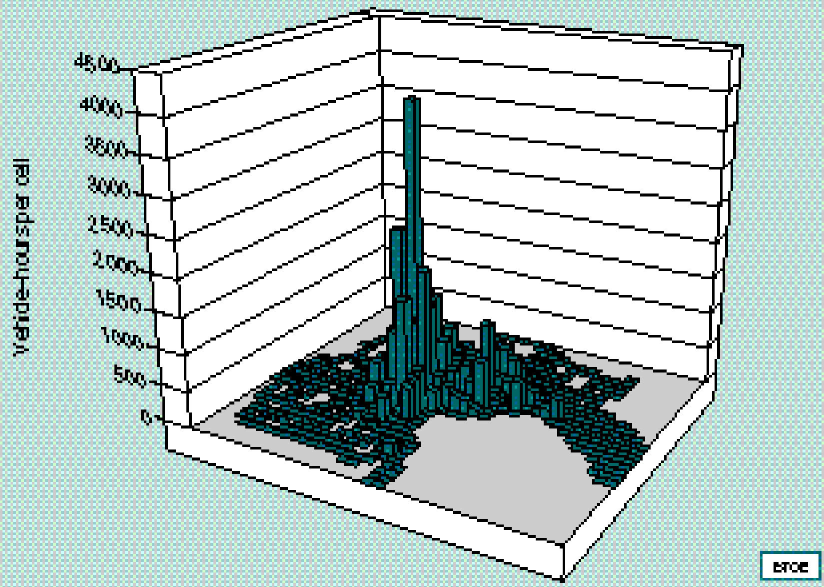
Source BTCE graphical representation of data from RJ Nairn & Partners (pers. comm.).

FIGURE 7.18 TRAVEL IN THE MELBOURNE AM PEAK HOUR



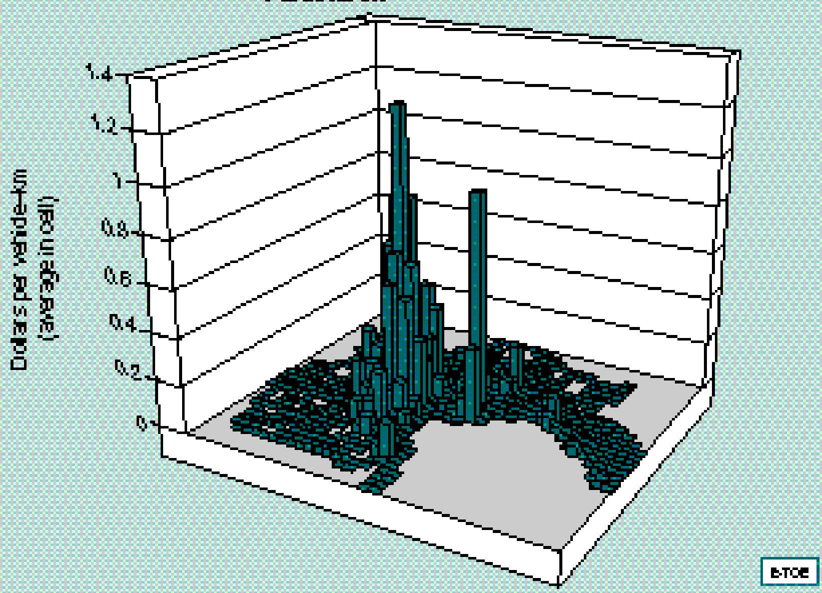
Source BTCE modeling results.

FIGURE 7.19 DELAY IN THE MELBOURNE AM PEAK HOUR



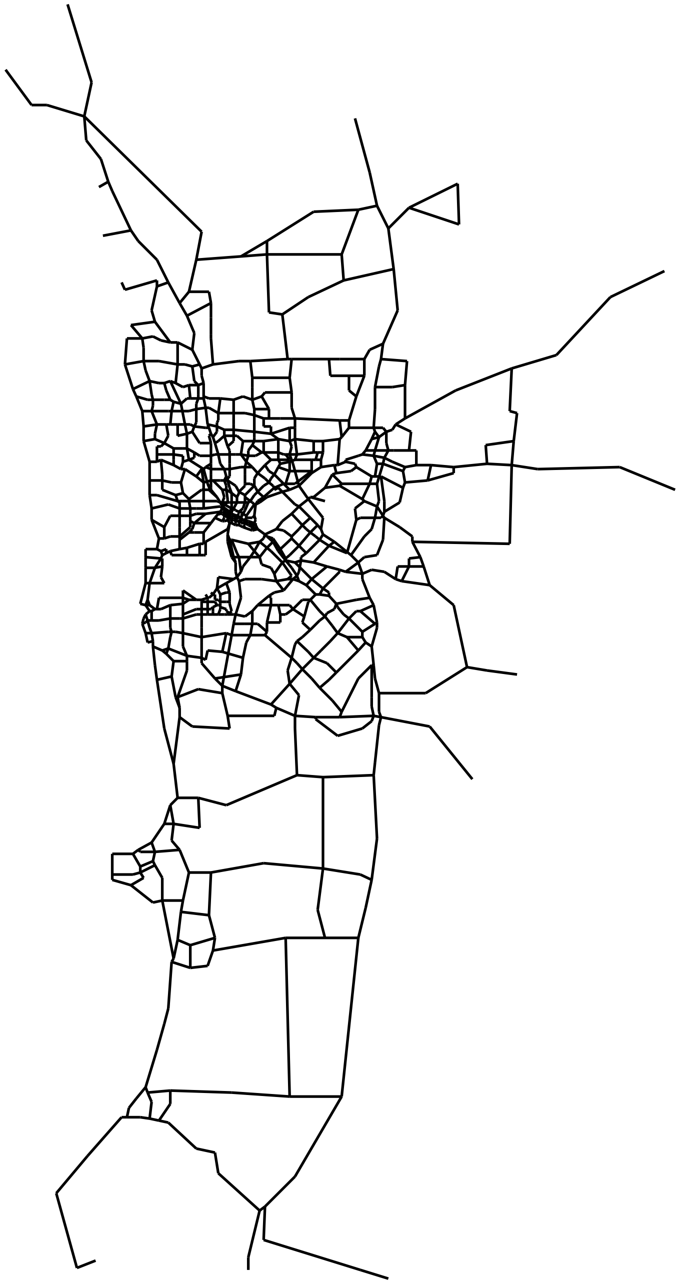
Source BTCE modelling results.

FIGURE 7.20 OPTIMAL ROAD USER CHARGES IN THE MELBOURNE AM PEAK HOUR



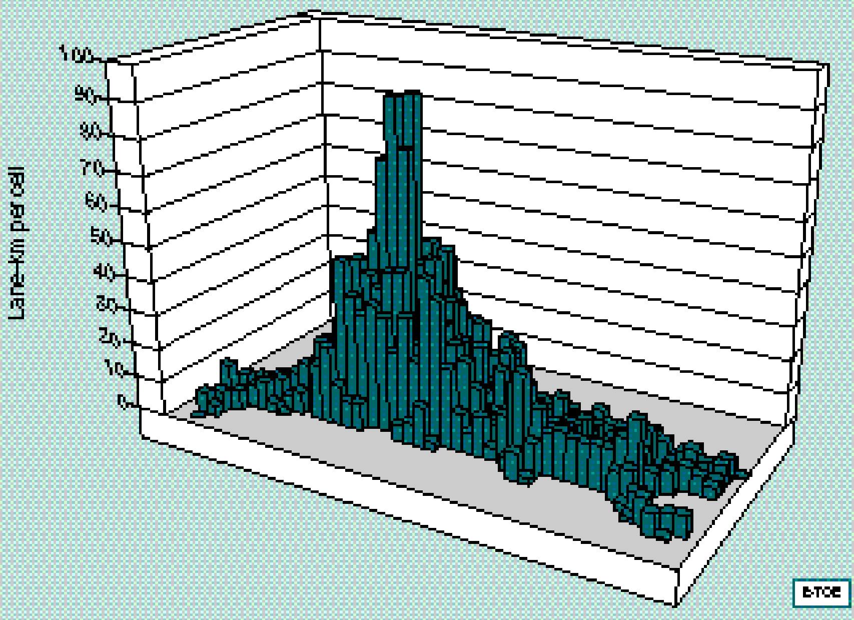
Source BTCE modelling results.

FIGURE 7.21 ROAD NETWORK USED IN THE PERTH MODEL



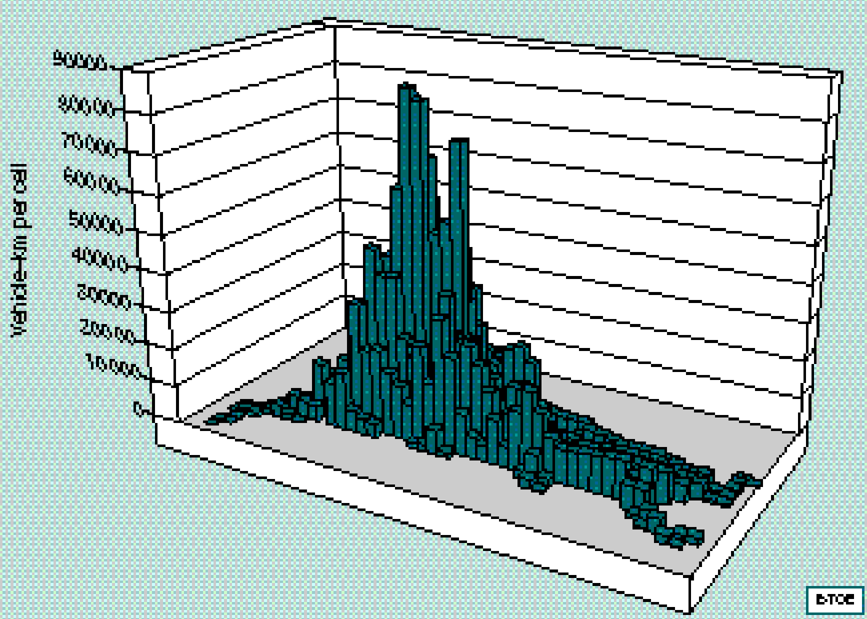
Source BTCE graphical representation of data from RJ Nairn & Partners (pers. comm.)

FIGURE 7.22 ROAD CAPACITY IN THE PERTH MODEL



Source BTCE graphical representation of data from RJ Nairn & Partners (pers. comm.).

FIGURE 7.23 TRAVEL IN THE PERTH AM PEAK HOUR



Source BTCE modelling results.

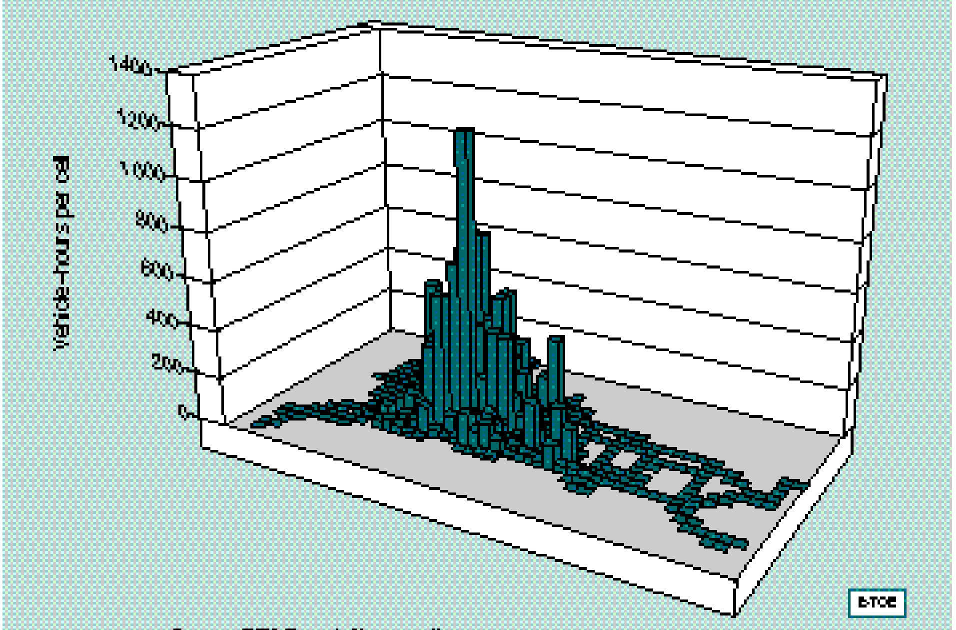


FIGURE 7.24 DELAY IN THE PERTH AM PEAK HOUR

Source BTCE modelling results.

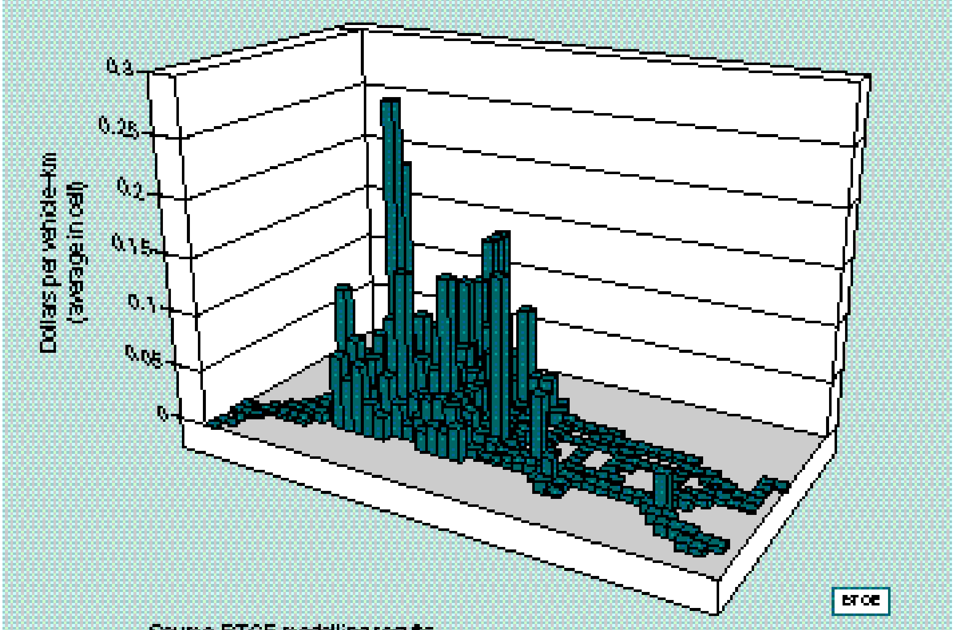


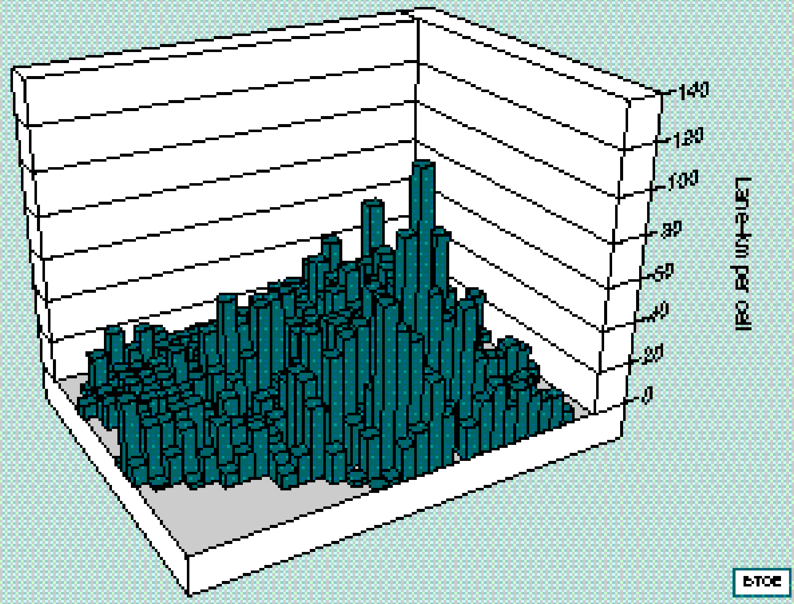
FIGURE 7.25 OPTIMAL ROAD USER CHARGES IN THE AM PERTH PEAK HOUR

Source BTCE modelling results.

FIGURE 7.26 ROAD NETWORK USED IN THE SYDNEY MODEL

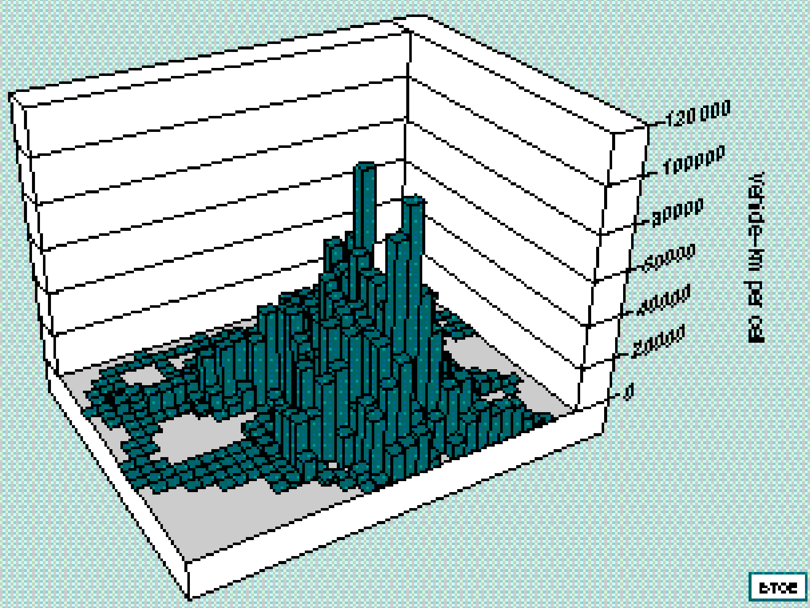
Source BTCE graphical representation of data from RJ Nairn & Partners (pers. comm.)

FIGURE 7.27 ROAD CAPACITY IN THE SYDNEY MODEL



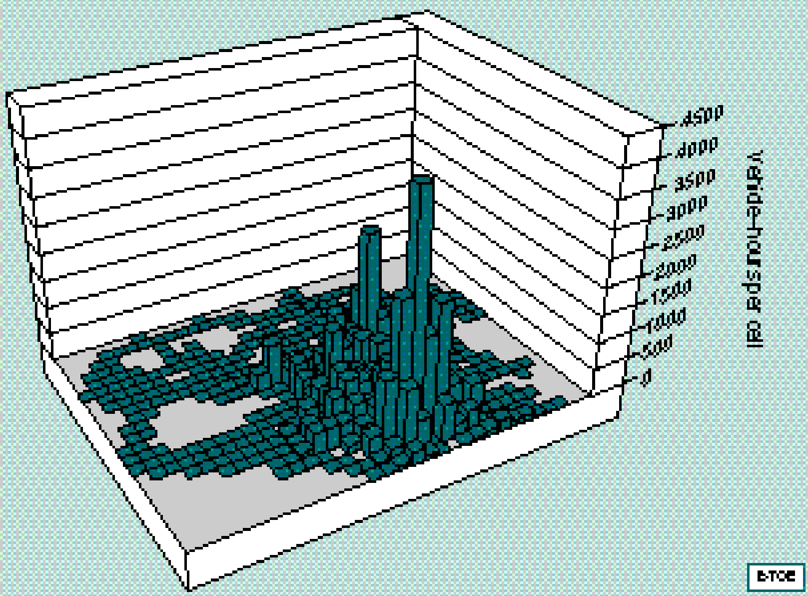
Source BTCE graphical representation of data from RJ Nairn & Partners (pers. comm.).

FIGURE 7.28 TRAVEL IN THE SYDNEY AM PEAK HOUR



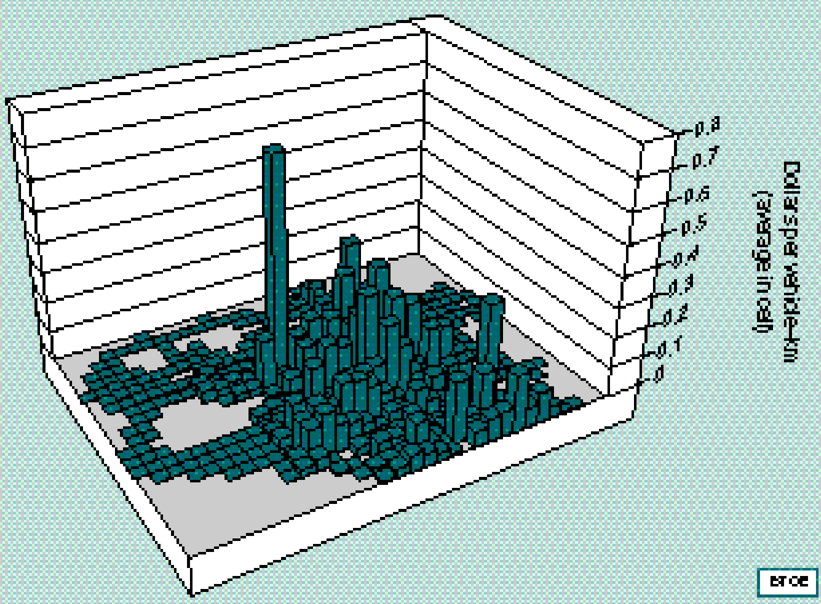
Source BTCE modelling results.

FIGURE 7.29 DELAY IN THE SYDNEY AM PEAK HOUR



Source BTCE modelling results.

FIGURE 7.30 OPTIMAL ROAD USER CHARGES IN THE SYDNEY AM PEAK HOUR



Source BTCE modelling results.

CHAPTER 8 CONCLUSIONS

It is important in drawing conclusions from this report to bear in mind that the analysis was an exploratory one. In particular, it was limited by the fact that it dealt only with commuter travel (journey to work) in the morning peak hour. Further, the data sets excluded commercial vehicles, whose continued growth in activity is likely to add significantly to urban congestion in the future.

More importantly, the charges estimated for sections of each city are based on data that range from the late 1980s to 1993. Developments since the collection of data (such as the completion of the Sydney harbour tunnel) are likely to affect these estimates significantly. The charges presented in this report should therefore be treated as being indicative only. The emphasis is on the methodology rather than on precise levels of charges.

A serious limitation when considering long-term impacts is that the model did not allow changes in travel cost and accessibility to feed back to the populations of the zones. Effectively, the model assumes no net movement of population between the zones in the city. It assumes that commuters may changed their residences to allow cheaper trips, but that emigration from any zone by commuters whose work is distant will be matched by immigration of commuters whose work is nearby and whose former homes were distant.

Finally, because TRANSTEP was not developed specifically as a behavioural model, a complete explanation of the results is not possible. For example, part of the reduction in peak travel under the modelled congestion charges may be a result of increased peak spreading, as motorists avoid congestion charges that are levied during peak travel times. Other explanations might include relocation or changing of jobs, relocation of employers to less congested areas, changes of school where children are dropped off by parents, etc. To clarify these explanations would require more research, preferably by linking TRANSTEP with a behavioural model.

Despite these limitations, the analysis undertaken has two main advantages over similar but more aggregated analyses employed by a number of other researchers. A major advantage is the degree of network detail used to estimate marginal costs of delay imposed by drivers on other commuters. Also, because the BTCE analysis has been carried out on a comparable basis for six major Australian capital cities, it affords a number of additional insights.

The comparison in chapter 5 between the detailed analysis of this work and various aggregated analyses shows two important results. First, there are significant shortcomings in even the best aggregated analysis. Second, there are significant advantages over even the best uniform road user charges to be gained by varying the level of road user charges according to local congestion. That differentiated charges should be more efficient has long been known. The magnitude of their superiority, as estimated in this work, suggests that it might be worth investigating the feasibility and cost of imposing a scheme of differentiated congestion charges.

Chapter 7 presents modelling results that illustrate the differences in patterns of traffic congestion between the cities studied. Clearly, there are major differences.

As might be expected, both the total number of trips and the kilometres travelled overall within each city fall as a result of congestion charges. It is also not surprising that the falls are smaller in the less populated cities: Adelaide, Canberra and Perth. Gains in average travel speed, reductions in travel times, and fuel savings are greatest in the more congested cities with initially lower average travel speeds: Sydney, Melbourne and Brisbane.

As is perhaps less obvious, the relative scales of various impacts (such as fall in travel and rise in speed) show major differences, which seem to result from the extent, rather than the intensity, of congestion.

In Melbourne, for example, congestion is concentrated on a relatively small central area near the CBD, where the economically efficient charge would be in the order of $1.26 per kilometre travelled. Only nine kilometres from the centre the charges are less than a tenth of this value. These concentrated charges encourage drivers to divert their routes around the central area. Thus average speeds over the whole city can be increased by 35 per cent at a cost of reducing road travel by a relatively modest 16 per cent.

In Sydney, on the other hand, high levels of congestion are more widespread, and are indeed ubiquitous from the CBD to the Parramatta area. Peak congestion charges would be lower (in the order of $0.75 per kilometre travelled at most), but significant charges would be imposed over a greater area. As a result, diversion to avoid tolls and congestion would not be so easy as in Melbourne, and the impacts would be rather different. Optimal congestion charges in Sydney would raise average speeds by 41 per cent, but at a cost of reducing travel by 27 per cent.

The ‘optimal’ road user charges described in this report are not the only possible treatment for the congestion problem in Australian capital cities. Nor are they necessarily the best treatment. For example, increasing road capacity might be a preferable alternative, or a useful adjunct, in some areas. It is even possible that the cost of collecting different charges on different roads at different times might erode their advantages. The results in chapter 7 simply show what the implications would be if it were decided to adopt an optimal pricing strategy to abate congestion in Australian capital cities.

APPENDIX I THE EFFECT OF TRAFFIC VOLUME ON SPEED

The essential feature of congestion is that, as traffic volumes increase, speeds decrease. The details of the relationship between speed and traffic flow are important for estimates of the costs of congestion, and even more so for estimates of marginal delay and congestion tolls. Unfortunately, the empirical basis for the relationship is weak.

Akçelik (1991) has provided illustrative functions (shown in figure I.1) which describe the speed–flow relationship for a number of broadly defined urban road types.

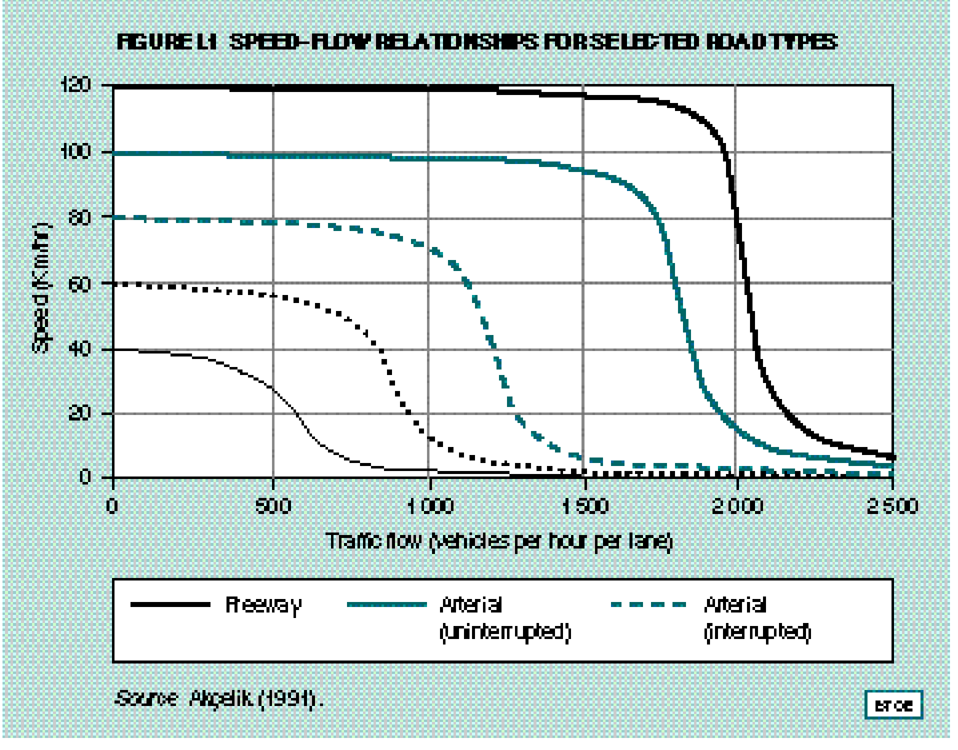


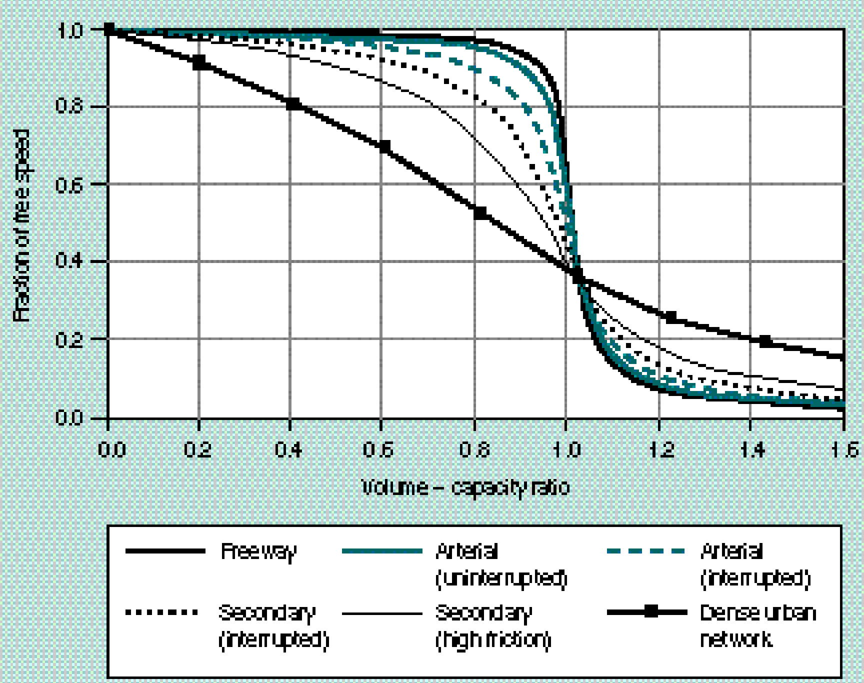
FIGURE I.1 SPEED–FLOW RELATIONSHIPS FOR SELECTED ROAD TYPES

Source Akçelik (1991).

For each of the road types in figure I.1, a free speed and a traffic flow capacity are defined. The interrupted arterial, for example, has a free speed of 80 kilometres per hour and a capacity of about 1200 vehicles per hour per lane. In addition, the shape of each curve is important. This is brought out more clearly by plotting the speed–flow relationship in *normalised* form, with speed measured as a fraction of the free speed and traffic represented by the volume-capacity ratio. Figure I.2 shows the speed–flow relationships of figure I.1 in normalised form, together with an area speed–flow relationship for a dense urban network.

The shape of the curves in figure I.2 changes progressively with road type. This is because there is a progressive change in the mechanism by which high traffic volumes reduce speed. At one extreme, on freeways, which are designed to provide minimal interference to traffic flow, the mechanism is mainly the interaction between vehicles within the traffic stream. Speed falls very little until traffic flow approaches capacity, after which it falls off very rapidly.

FIGURE I.2 NORMALISED SPEED–FLOW RELATIONSHIPS FOR SELECTED ROAD TYPES



Source Akçelik (1991) and BTCE estimates.

At the other extreme, roads forming part of a dense urban network are continually interrupted by intersections and other discontinuities. On these roads, speeds are reduced primarily because of the time spent in queues at intersections and discontinuities. Speeds start to fall at low traffic volumes and fall continuously to low values. There is no very obvious critical value at which speeds fall rapidly, and it is hard, therefore, to identify capacity.

The curve for the dense urban network is included to illustrate the limiting case of this progression of road types, though it is not strictly part of the same series as the other curves in figure I.2. The dense urban network curve is a semi-empirical result derived from modelling work in Hong Kong (Harrison et al, 1986). It differs from the other curves because it describes the average speed of traffic in an area rather than on a single link. There is also a normalisation problem associated with the difficulty in identifying capacity. The scale of the normalised curve in the horizontal direction in figure I.2 is, therefore, uncertain. It has been chosen so that the curve passes close to the point of intersection of Akçelik's curves, in order to highlight the progressive change in shape.

The analysis of congested traffic flow and attempts to derive mathematical expressions for the form of these curves (Davidson 1978; Akçelik 1978; Taylor 1984; Tisato 1991) have, in fact, focussed on travel time, the inverse of speed. The most thorough treatment appears to be due to Kimber and Hollis (1979), who derived a function which correctly describes travel time and intersection delay on a road link for values of traffic flow both above and below capacity. Their function takes into account random fluctuations in traffic flow below capacity, the effect of overcapacity traffic flows of finite duration, and provides a smooth transition between the two regimes.

Kimber and Hollis's function has the form

$$t_a = t_o \left[1 + a \left\{ (x-1) + \sqrt{(x-1)^2 + bx} \right\} \right] \qquad \text{(I.1)}$$

where

t_a = average travel time per kilometre;

t_o = free speed travel time per kilometre;

x = volume–capacity ratio; and

a and b are parameters.

Akçelik's normalised speed–flow curves plotted in figure I.2 are obtained from this equation by plotting $\frac{t_o}{t_a}$ against x for the following values of a and b:

Freeway	a = 30;	b = 0.00040
Arterial (uninterrupted)	a = 25;	b = 0.00089
Arterial (interrupted)	a = 20;	b = 0.00267
Secondary (interrupted)	a = 25;	b = 0.00711
Secondary (high friction)	a = 10;	b = 0.02133

The curve for the dense urban network can be closely approximated by $a = 4$; $b = 0.17$.

It is possible to interpret the parameters a and b in terms of the traffic engineering quantities as discussed by Kimber and Hollis (1979) and by Akçelik (1991). However, for this work, it is sufficient that the function is simple, differentiable and defined for all positive x's.

The choice of parameters appropriate for this work is complex. Because of the limited empirical basis, it has been necessary to experiment with ranges of parameter values. And, because some of the results depend strongly on these values, it has been necessary to test the sensitivity to alternative choices. Fortunately, the choice is not unconstrained. For example, attempts to calibrate the model with parameter values close to those of figure I.1 lead to implausibly high average speeds. This can be rectified by reducing free speeds or capacities or changing the speed–flow relationships. Details of the calibration and sensitivity testing are described in chapter 6.

Table I.1 gives the parameter values on which the results reported here are based, and figure I.3 shows the normalised speed–flow relationships for some of the road types marked in table I.3.

The economic analysis also requires the marginal travel time function. This is the derivative, with respect to traffic, of the total travel time. If

T = total travel time

q = traffic volume

Q = capacity of the link ($x = q/Q$)

then

$$T = qt_a$$

and the marginal travel time function has the form

$$t_m = \frac{dT}{dq} = q\frac{dt_a}{dq} + t_a \tag{I.2}$$

where

$$\frac{dt_a}{dq} = t_o a\left[1 + \frac{(x-1) + b/2}{\sqrt{(x-1)^2 + bx}}\right]$$

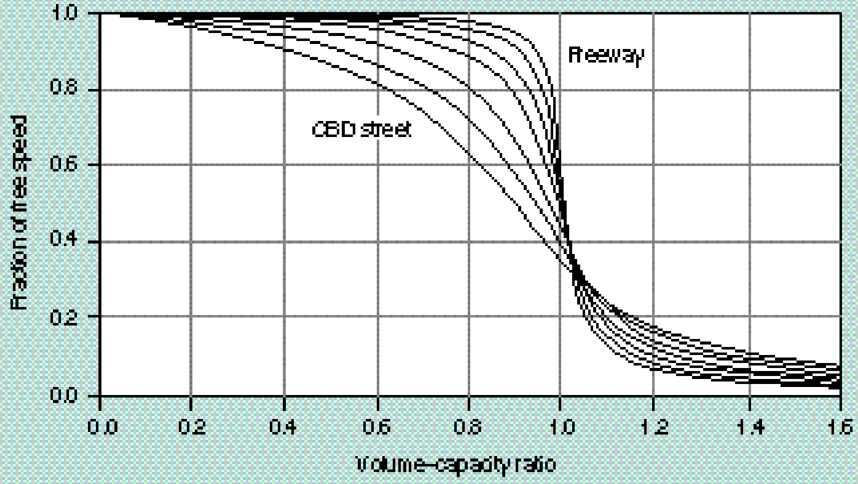
TABLE I.1 PARAMETERS OF THE SPEED–FLOW RELATIONSHIP FOR ROAD TYPES IN THE TRANSTEP MODEL

Road type	*Road capacity (veh/hr/lane)*	*Free speed (km/hr)*	*Speed–flow parameters*	
			a	*b*
Urban freeway high standard[a]	2040	100.0	31.26	0.0004
Urban tollway — 4 or more lanes	1980	90.0	28.00	0.0006
Urban expressway[a]	1920	80.0	25.07	0.0009
Urban expressway — frequent signals	1860	70.0	21.82	0.0015
Divided arterial — clear running[a]	1800	60.0	20.18	0.0020
Divided arterial — no parking — signals	1740	50.0	18.60	0.0027
Divided arterial — parking[a]	1560	45.0	16.71	0.0040
Undivided arterial — clearway[a]	1500	42.5	13.02	0.0100
Undivided arterial — trams — parking	1380	40.0	11.66	0.0150
Distributor — bus route[a]	1320	37.5	10.79	0.0200
Local road — collector	1260	35.0	10.15	0.0250
CBD street[a]	1200	30.0	9.26	0.0350

a. Plotted in figure I.3.

Source R. J. Nairn and Partners and BTCE.

FIGURE I.3 NORMALISED SPEED-FLOW RELATIONSHIPS FOR ROAD TYPES (MARKED IN TABLE I.3) IN THE TRANSTEP MODEL



Source BTCE estimates.

APPENDIX II ESTIMATES OF ECONOMIC MEASURES OF CONGESTION FOR THE BUSINESS COUNCIL OF AUSTRALIA

Meyrick (1994) used data from Commeigne's (1992) study of the cost of congestion in Sydney to estimate the toll, revenue and net benefit from imposing a uniform congestion toll over the whole of the Sydney metropolitan area.

Commeigne's basic assumption was that average traffic speeds in Sydney were reduced by congestion from 45 km / hr to 40 km / hr, and his study was principally concerned to estimate the value of the resulting increases in travel time (delay) and vehicle operating costs. Commeigne was not concerned with congestion tolls.

Meyrick's analysis assumed:

- a linear speed–flow function of the form

$$S = S_{max} - S_1 \quad V$$

  where S is the average speed of travel on the whole network, S_{max} is the average free speed on the network (that is the average speed if there were no congestion), S_1 is a coefficient describing how the average speed is reduced by the quantity of travel, and V is the annual quantity of travel on the whole network;

- average travel speed (S in 1992) of 40 km / hr, reduced by congestion from an average free speed (S_{max}) of 45 km / hr;

- total annual travel (V in 1992) on the whole network of 25.3 billion veh-km;

- speed dependent vehicle operating costs (*VOC*) of the form

$$VOC = a + \frac{b}{S}$$

where a = $0.18 / veh-km and b = $4.27 / veh-hr;

- an average value of time T of $20 / veh-hr; and
- an elasticity of demand for travel of -0.2.

With these assumptions, Meyrick obtained:

- an optimal uniform congestion toll of $0.076 / veh-km;
- annual revenue from this toll of $1.92 billion; and
- annual net benefit from the imposition of the toll of $20 million, or about 1 per cent of the toll revenue.

Using the above notation, the average cost of travel is

$$a + \frac{b + T}{S_{max} + S_1 \quad V},$$

and the marginal cost of travel is

$$a + \frac{S_{max} \quad (b + T)}{(S_{max} + S_1 \quad V)^2}$$

Figure II.1 shows these cost curves (drawn to scale) together with a linear demand curve which has an arc elasticity of -0.2 between its intersections with the cost curves. The form of the curves provides some insight into Meyrick's results, particularly the very small value of net benefit—represented by the very small triangular area to the right of the demand curve (area ABE in figure 4.1).

The results are dominated by the choice of the linear speed–flow relationship and the assumption that the current quantity of travel is far from any overall capacity constraint that could lead to sharply rising average costs. As a result, the external cost (the difference between the marginal and average cost) is small. Together with the steep slope of the low elasticity demand curve, this leads to the very small net benefit triangle.

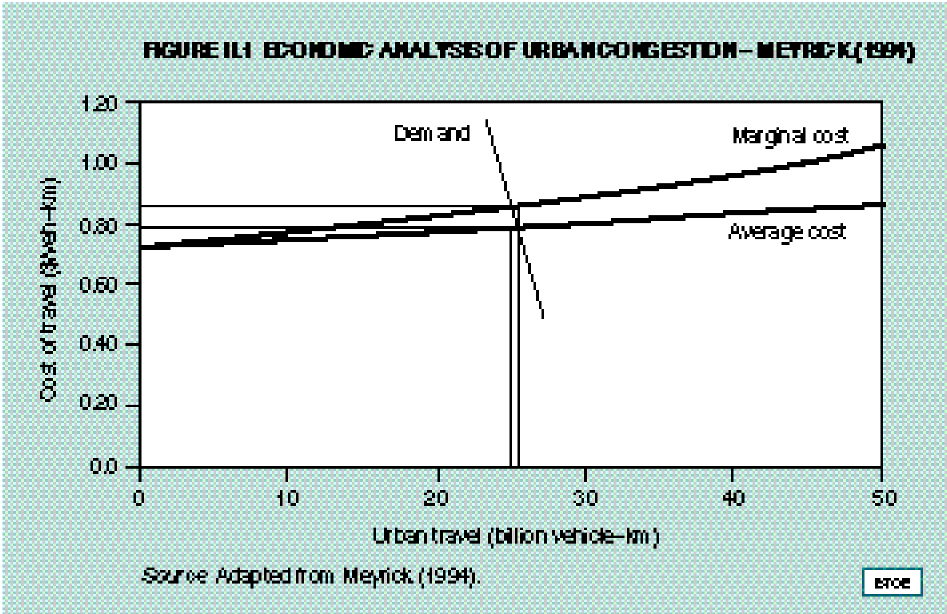


FIGURE II.1 ECONOMIC ANALYSIS OF URBAN CONGESTION – MEYRICK (1994)

Source Adapted from Meyrick (1994).

APPENDIX III THE ARRB MODEL OF CONGESTION TOLLS AND REVENUES

The ARRB estimates of congestion tolls and revenues in Australian cities were obtained using the ARRB Travel Cost Model (ATCM) developed by Hepburn and Luk (1994). The version used for the work reported in this paper is known as model P.

The ATCM is a simple spreadsheet model which deals with urban traffic flow on five road types and during four periods on the typical weekday.

The road types are:

- freeways;
- CBD streets;
- arterial (inner)—(for example within 15 km of the CBD in the case of Melbourne);
- arterial (outer); and
- local roads.

The typical weekday is divided into the following periods:

- the morning peak;
- the inter-peak period;
- the evening peak; and
- off-peak periods.

Annual traffic flow (in vehicle-kilometres) in each city is obtained from the Australian Bureau of Statistics Survey of Motor Vehicle Usage (SMVU) and divided by 300 to convert it to a daily flow. The daily flow is allocated independently among the road types and periods (in other words, the distribution of traffic over the road types is the same in each period). The distributions are similar but not identical in each city, as shown in table III.1.

The basic unit of analysis in the model is the traffic flow on one of the road types during a particular period of the day. This analysis produces a set of results which are specific to each combination of road type and period (CBD during the morning peak, or arterial (inner) roads during the off-peak period). The specific results are:

- the level of the theoretically optimal toll;
- the percentage reduction in traffic level as a result of applying the optimal toll; and
- the percentage increase in travel speed as a result of applying the optimal toll.

These results can be combined with the traffic flow to give aggregate economic measures for the traffic using that road type in that period:

- the toll revenue;
- the consumer surplus losses by motorists who refuse to pay the toll (the tolled-off);

TABLE III.1 ANNUAL AND DAILY TRAFFIC AND ITS DISTRIBUTION AMONG ROAD TYPES AND TRAFFIC PERIODS

	Melbourne	*Sydney*	*Brisbane*	*Adelaide*	*Perth*
Annual traffic (‘000 vkt)	23 952 600	24 952 500	11 082 300	8 141 400	10 125 000
Daily traffic (‘000 vkt)	79 842	83 175	36 941	27 138	33 750
Distribution (road type)					
Freeways	0.07	0.07	0.07	0.00	0.07
CBD streets	0.02	0.03	0.02	0.04	0.02
Arterial (inner)	0.33	0.325	0.33	0.355	0.33
Arterial (outer)	0.33	0.325	0.33	0.355	0.33
Local	0.25	0.25	0.25	0.25	0.25
Distribution (traffic period)					
AM Peak	0.20	0.20	0.18	0.15	0.15
Inter-Peak	0.35	0.35	0.35	0.35	0.35
PM Peak	0.30	0.30	0.17	0.15	0.15
Off-Peak	0.15	0.15	0.30	0.35	0.35

Source ARRB Contract Report No. CR TE 95/012.

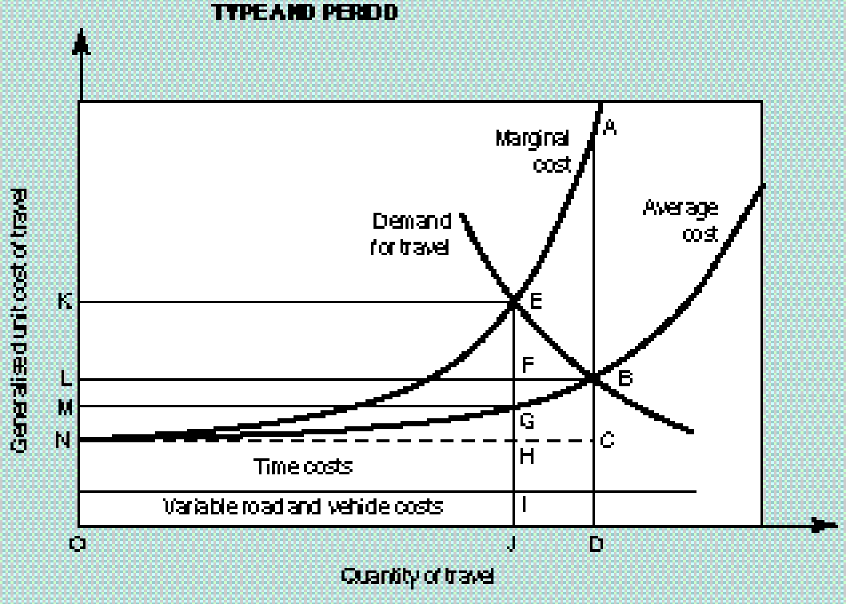
- the consumer surplus losses by motorists who remain and pay the toll; and
- the net benefit to society from charging the toll.

These aggregates can be combined to give daily and annual totals.

The essential assumptions and the sequence of calculations can be described by referring to the standard diagram for the economic analysis of congestion, figure III.1

The marginal and average cost curves describe how delay costs increase with the volume of traffic. They are based on a family of curves reported by Akçelik (1991). For each road type a free speed (corresponding to low traffic volumes) and a congested speed (corresponding to traffic volumes near capacity) are provided, as shown in table III.2.

FIGURE III.1 CONGESTION ANALYSIS FOR A PARTICULAR ROAD TYPE AND PERIOD



The current level of traffic (for a road type and period) is positioned with respect to the marginal and average cost curves by estimating the volume–capacity ratio under which it operates. This establishes the point D and the corresponding points A and B on the cost curves. The volume–capacity ratio on all roads during the off-peak and on local roads at all times is assumed to be 0.5. The values assumed for the other roads and periods are given in table III.3.

A value of travel time is used to relate congestion delay to other costs. For all except local roads its value is $12 per vehicle–hour during the peak periods and $16 per vehicle–hour during the inter-peak period. A value of $1 per vehicle-hour is used for all road types during the off-peak and for local roads at all times.

TABLE III.2 FREE SPEED AND CONGESTED SPEED ASSUMED FOR EACH CITY AND ROAD TYPE[a]

(kilometres per hour)

Road type	*Melbourne*	*Sydney*	*Brisbane, Adelaide & Perth*
Freeways	87/55	80/54	79/52
CBD streets	46/20	45/19	36/22
Arterial (inner)	64/40	60/38	53/39
Arterial (outer)	64/40	60/38	53/39
Local	46/20	45/19	36/22

a. Entries given as free speed/congested speed.

Source ARRB under contract to BTCE (Report CR TE 95/012).

TABLE III.3 VOLUME–CAPACITY RATIOS ASSUMED FOR CURRENT TRAFFIC BY ROAD TYPE, PERIOD AND CITY

		Inter-peak period		
Road type	*Peak period All cities*	*Melbourne & Sydney*	*Brisbane*	*Adelelaide & Perth*
Freeway	1.0	0.8	0.8	0.7
CBD	1.0	0.9	0.8	0.7
Arterial (inner)	1.0	0.9	0.8	0.7
Arterial (outer)	0.9	0.8	0.8	0.7

Source ARRB under contract to BTCE (Report. CR TE 95/012).

The demand curve is assumed to have constant elasticity. Elasticities for inter-peak and off-peak periods were set at -0.4 and -0.5 respectively. The sensitivity of the results to changes in the peak period elasticities is tested over the range -0.2 to -1.2.

A value is assumed for the variable costs of travel. These are primarily fuel, maintenance and vehicle operating costs. They are the costs as perceived by the motorist, and therefore include any taxes and excise. The variable costs of travel are assumed to be independent of congestion, and therefore only affect the vertical position of the cost curves. A single value of 10 cents/km is assumed for all road types and periods.

With this set of assumptions, the sequence of calculations proceeds as follows:

- the assumed volume–capacity ratio (table III.3) establishes the points A, B and D;
- the demand curve with the assumed constant elasticity is scaled so that it passes through B, its intersection with the marginal cost curve establishing the point E and the corresponding volume–capacity ratio J;
- the results (for each road type and period) follow
  - the level of the theoretically optimal toll as given by GE;
  - the percentage reduction in traffic level as a result of applying the optimal toll is given by 100*JD/OD;
  - the percentage increase in travel speed as a result of applying the optimal toll is given by 100*(IF-IG)/IG;
  - the toll revenue is given by the area KEGM;
  - the consumer surplus lost by motorists who refuse to pay the toll (the tolled-off) is given by the area EBF;
  - the consumer surplus lost by motorists who remain and pay the toll is given by the area KEFL;
  - the net benefit to society from charging the toll (the revenue minus the consumer surplus losses) is given by the area ABE; and
  - the cost of congestion (defined as the value of the difference in travel time under current conditions and under free flow conditions) is given by the area LBCN; and

- these can then be combined to give daily and annual totals.

RESULTS

Aggregate economic measures of congestion

Table III.4 gives the main economic measures of congestion for Melbourne, Sydney, Brisbane, Adelaide and Perth for an elasticity of travel demand of -0.4.

Figure III.2 shows the sensitivity of these results to different assumptions about demand elasticity. For each city, a series of six bars is presented which record annual value estimates in $billion per year.

The first bar for each city shows the annual *cost of congestion* (area LBCN in figure III.1).

The cost of congestion does not depend on assumptions about the elasticity of travel demand.

The other five bars for each city show the revenue, consumer surplus losses and net benefit from imposing optimal congestion tolls. Estimates of these quantities depend on assumptions about the elasticity of travel demand. The five bars show the sensitivity of these estimates to variation of the assumed peak period demand elasticity in the range -0.2 to -1.2.

TABLE III.4 ECONOMIC MEASURES OF CONGESTION IN AUSTRALIAN CITIES[a]

($million per year)

	Melbourne	*Sydney*	*Brisbane*	*Adelaide*	*Perth*
Cost of congestion[b]	809	926	176	111	133
Congestion toll measures[c]					
Toll revenue	1 608	1 816	384	230	275
Road user losses[d]	1 141	1 295	278	164	191
Net benefit[e]	466	521	107	67	83

a. Summed over all road types (except local) and all traffic periods (except off-peak).
b. The value of time lost due to congestion.
c. As compared with the current congested situation with no toll.
d. Losses to all road users, the tolled and the tolled-off.
e. Net benefit = toll revenue –road user losses.

Source ARRB under contract to BTCE (Report No. CR TE 95/012).

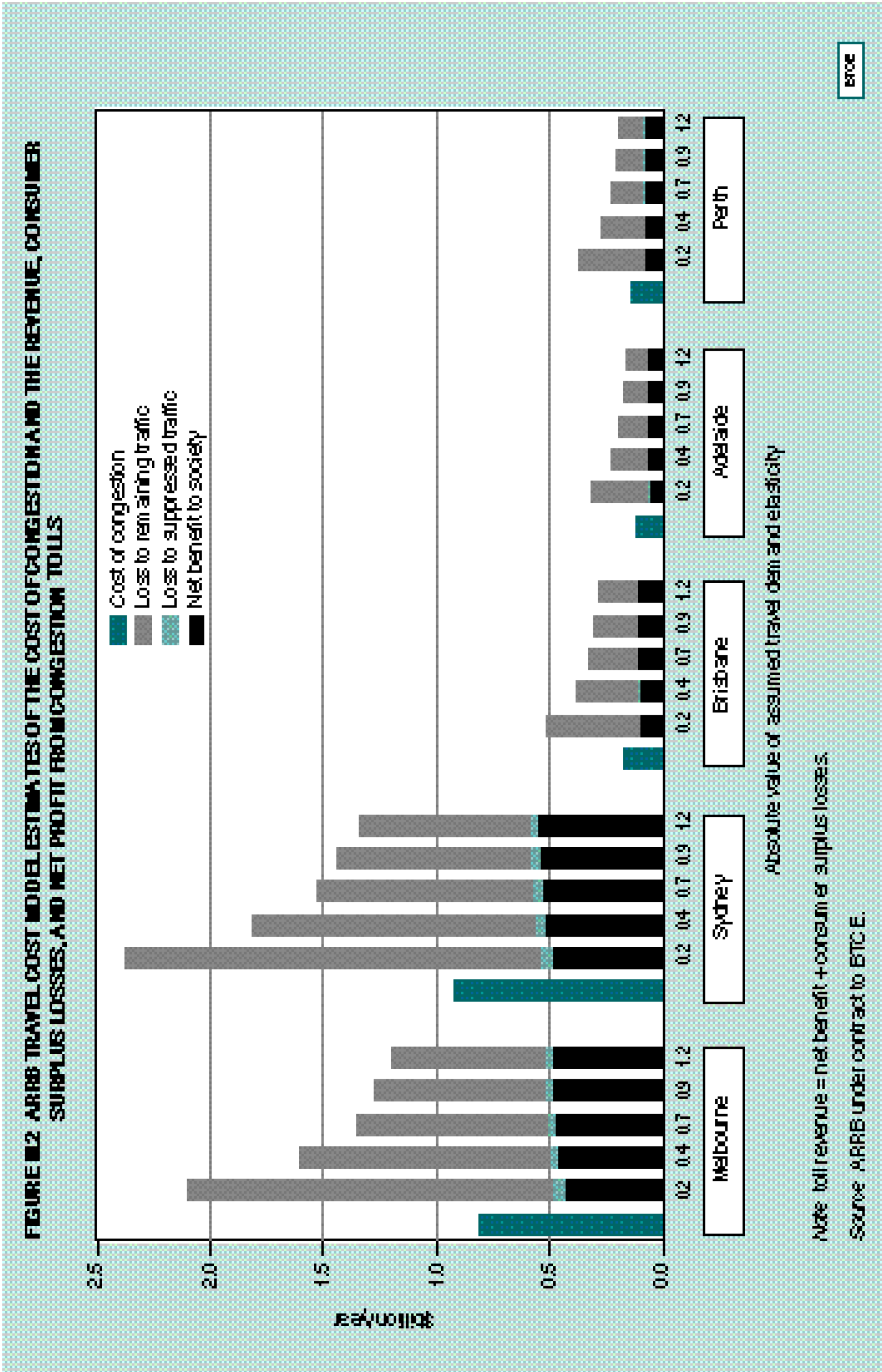


FIGURE III.2 ARRB TRAVEL COST MODEL ESTIMATES OF THE COST OF CONGESTION AND THE REVENUE, CONSUMER SURPLUS LOSSES, AND NET PROFIT FROM CONGESTION TOLLS

Note: toll revenue = net benefit + consumer surplus losses.

Source: ARRB under contract to BTCE.

The full height of the bar represents the *revenue* obtained by imposing the toll (area KEGM in figure III.1).

The upper section of the bar represents the *consumer surplus lost by remaining motorists* who pay the toll (area KEFL in figure III.1).

The small middle section of the bar represents *the consumer surplus lost by motorists who are 'tolled off'*—those for whom the costs of travel, including the toll, now exceed the benefits (area EBF in figure III.1).

The lower section of the bar represents the *net benefit* to society from imposing the toll (area ABE in figure III.1).

The sensitivity of these results to varying assumptions about the demand elasticity is shown. Long run elasticities of demand for urban car travel with respect to petrol price appear to be in the range -0.2 to -0.4. However, the relevance of these elasticities to motorists' response to congestion tolls is an open question.

Congestion charges and the impact on traffic conditions

The optimal congestion toll varies with traffic conditions and is therefore different on different road types and at different times. Tables III.5 and III.6 present the toll levels and the consequences for traffic conditions during the peak and inter-peak periods. The first section of table III.5 shows the optimal toll (in cents per veh-km) during the morning and evening peak periods (which have similar traffic conditions) for four road types in Melbourne, Sydney, Brisbane, Adelaide and Perth.

Table III.5 also shows the impact of these tolls on traffic conditions in terms of the percentage reduction in traffic volumes and the percentage increase in speeds.

Table III.6 presents the corresponding results for the inter-peak period.

TABLE III.5 CONGESTION TOLLS AND CONSEQUENT CHANGES IN TRAFFIC CONDITIONS — PEAK PERIOD

	Melbourne	Sydney	Brisbane	Adelaide	Perth
Optimal toll (cents/km)					
Freeways	14	13	14	..	14
CBD streets	57	62	40	40	40
Arterial (inner)	20	21	16	16	16
Arterial (outer)	7	7	5	5	5
Traffic reduction (%)					
Freeways	8	8	8	..	8
CBD streets	15	15	12	12	12
Arterial (inner)	9	9	8	8	8
Arterial (outer)	7	7	5	5	5
Speed increase (%)					
Freeways	46	39	41	..	41
CBD streets	66	67	41	41	41
Arterial (inner)	45	43	29	29	29
Arterial (outer)	4	4	1	1	1

.. Not applicable.

Source ARRB under contract to BTCE.

TABLE III.6 CONGESTION TOLLS AND CONSEQUENT CHANGES IN TRAFFIC CONDITIONS — INTER-PEAK PERIOD

	Melbourne	Sydney	Brisbane	Adelaide	Perth
Optimal toll (cents/km)					
Freeways	2	2	2	..	1
CBD streets	37	40	11	6	6
Arterial (inner)	9	9	2	1	1
Arterial (outer)	4	4	2	1	1
Traffic reduction (%)					
Freeways	3	2	3	..	1
CBD streets	13	14	6	4	4
Arterial (inner)	7	7	2	1	1
Arterial (outer)	4	4	2	1	1
Speed increase (%)					
Freeways	0	0	0	..	0
CBD streets	23	25	2	1	1
Arterial (inner)	4	4	0	0	0
Arterial (outer)	1	1	0	0	0

.. Not applicable.

Source ARRB under contract to BTCE.

REFERENCES

ARRB	Australian Road Research Board
BCA	Business Council of Australia
BTCE	Bureau of Transport and Communications Economics
TRRL	Transport Road Research Laboratory

Akçelik, R. 1978, 'A new look at Davidson's travel time function' *Traffic Engineering and Control*, vol. 19, pp. 459–463.

——1991, 'Travel time functions for transport planning purposes: Davidson's Function, its time-dependent form and an alternative travel time function' *Australian Road Research*, vol. 21, no. 3, pp. 49–59.

BTCE 1993, 'Goods and bads in urban transport' *Transport and Communications Indicators,* Bulletin 42, September Quarter.

Commeignes, H. 1992, *The Cost of Congestion in Sydney*, a report prepared for the Roads and Traffic Authority of New South Wales, Pricing Strategy Branch, February 1992.

Davidson, K.B. 1978, 'The theoretical basis of a flow travel–time relationship for use in transportation planning' *Australian Road Research*, vol. 8, no. 1, pp. 32–35.

Else, P.K. 1981, 'A reformulation of the theory of optimal congestion taxes' *Journal of Transport Economics and Policy*, vol. 15, pp. 217–232.

—— 1982, 'A reformulation of the theory of optimal congestion taxes: A rejoinder' *Journal of Transport Economics and Policy*, vol. 16, pp. 299–304.

—— 1986, 'No entry for congestion taxes?' *Transportation Research-A*, vol. 20A, no. 2, pp. 99–107.

Evans, A.W. 1992, ‘Road congestion pricing: when is it a good policy?’ *Journal of Transport Economics and Policy*, September, pp. 213–243.

Harrison, W.J., Pell, C., Jones, P.M. & Ashton, H. 1986, ‘Some advances in model design developed for the practical assessment of road pricing in Hong Kong’ *Transportation Research-A*, vol. 20A, no. 2, pp. 135–143.

Hau, T.D. 1992, *Economic Fundamentals of Road Pricing A Diagrammatic Analysis*, World Bank Working Paper WPS 1070, The World Bank, Washington.

Hensher, D.A 1993, *Greenhouse Gas Emissions and the Demand for Urban Passenger Transport: Design of the Overall Approach*, Occasional Paper 108, BTCE, Canberra.

Hepburn, S. & Luk, J.Y.K. 1994, *User guide to the ARRB Travel Cost Model (ATCM)*, Working Document No. WD TE 94/001, ARRB, Melbourne.

Hills, P. 1993, ‘Road congestion pricing: when is it a good policy? A comment’ *Journal of Transport Economics and Policy*, vol. 27, pp. 91–99.

Kimber, R.M. & Hollis, E.M. 1979, *Traffic Queues and Delays at Road Junctions*, TRRL Laboratory Report 909, Transport and Road Research Laboratory, Berkshire, U K.

Luk, J. 1994, ‘An area model for the investigation of congestion pricing schemes’, Paper presented at the international seminar on *Transportation Planning and Policy in a Network and Price Equilibrium Framework*, 10–11 August 1994, Canberra.

Luk, J. & Hepburn, S. 1993, *A Review of Australian Travel Demand Elasticities*, Working Document No. TE 93/004, ARRB, Melbourne.

Meyrick, S.J. 1994, ‘Road user charges: Discussion paper’ (Annex A) in *Refocussing Road Reform* by J.B. Cox, a report prepared for the Business Council of Australia, BCA, Melbourne, pp. 229–296.

Nairn, R.J., Field, J.F. & Parker, G.R. 1977, ‘Land-use/transport interaction modelling with TRANSTEP’ *3rd Australian Transport Research Forum*, 24–25 May, Ministry of Transport, Melbourne, pp. 339–382.

Nairn & Partners (unpub.), General description of the TRANSTEP Model 1986.

—— (unpub.), TRANSTEP Model User Manual 1991.

Pigou, A.C. 1920, *The Economics of Welfare*, Macmillan, London.

Small, K.A. 1992, *Urban Transportation Economics*, Harwood, Chur, Switzerland.

Taylor, M.A.P. 1984, 'A note on using Davidson's Function in equilibrium assignment' *Transportation Research B*, vol. 18B, no. 3, pp. 181–199.

Tisato, P. 1991, 'Suggestions for an improved Davidson travel time function' *Australian Road Research*, vol. 21, no. 2, pp. 85–99.