Bureau of Transport Economics

WORKING PAPER 39

URBAN TRANSPORT MODELS: A REVIEW

The Bureau of Transport Economics undertakes applied economic research relevant to the portfolios of Transport and Regional Services, and Employment, Workplace Relations and Small Business. This research contributes to an improved understanding of the factors influencing the efficiency and growth of these sectors and the development of effective policies.

A list of recent publications appears on the inside back cover of this publication.

Bureau of Transport Economics

WORKING PAPER 39

URBAN TRANSPORT MODELS: A REVIEW

ISSN 1440-9707
ISBN 0 642 28305 2

This publication is available free of charge from the Bureau of Transport Economics, GPO Box 501, Canberra, ACT 2601, or by phone 02 6274 6846, fax 02 6274 6816 or email *bte@dot.gov.au*

http://www.dot.gov.au/programs/bte/btehome.htm

Printed by the Department of Transport and Regional Services

PREFACE

Transport models have been used for several decades now, both for research, and as an analytical tool to assist planners and decision-makers.

As the complexity of traffic and environmental problems in our cities has increased, policy makers have come to depend on models to an even greater extent. The immense increase in available computing power over the last decade has abetted this dependence. Customised software has simplified even the most complex mathematics to such an extent that modelling is no longer the preserve of a select few 'rocket scientists'.

If asked, however, many policy analysts and decision-makers would probably admit to a lack of understanding of the models on the results of which they rely. Billions of dollars in resources are expended annually in Australia despite a lack of full understanding of the basis on which decisions are made.

To assist both researchers and decision-makers, Dr William (Weiguo) Lu has dissected the major models that have been used to analyse urban transport tasks. This Working Paper therefore represents something of a 'scene-setter' for further work.

While a purely non-technical approach is not feasible, he has sought to provide an intuitive exposition of the basic concepts involved, relying on a minimum of mathematical expression.

The project was carried out under the supervision of Dr Mark Harvey. Comments on drafts were provided by Dr Leo Dobes and David Mitchell.

Dr Leo Dobes
Research Manager

Bureau of Transport Economics
Canberra
October 1998

CONTENTS

FIGURES

TABLES

BOXES

ABSTRACT

Over the last three decades, urban transport modelling has seen a gradual incorporation and unification of different theories and methods, resulting in a more consistent framework and reflecting a deeper understanding of the real world situation. The emergence of integrated land use–transport models incorporating behavioural relationships represents the culmination of efforts in recent years.

At least four broad types of model are used by researchers and planners: traditional four-step transport models; behavioural travel demand models; linked urban land use–transport models; and integrated urban land use–transport models.

Following a detailed analysis of the theoretical bases of each type of model, comparisons are made from a predictive and a policy analysis perspective. As expected, each type of model has its particular strengths and weaknesses. By their nature, however, integrated models offer a more comprehensive potential for analysing policy changes.

CHAPTER 1 OVERVIEW

Urban transport accounts for an estimated 68 per cent of the total vehicle-kilometres travelled (for both passengers and freight) in Australia (ABS 1995, p. 15; BTE estimate).

Transport planners and policy makers confront a much wider range of issues today than ever before. Planning and policy interests have now broadened to encompass safety, efficiency, equity, and environmental sustainability in addition to the traditional concerns of investment and congestion. The broadening of focus has increased the complexity of the planning and policy task very significantly.

In the United States, transport planning is required by legislation as a condition for the receipt of federal transport funds by larger urban areas. The most recent legislation, *The Intermodal Surface Transportation Efficiency Act,* has led to even stricter regulations, requiring planning authorities to deal with a wide range of issues, including air quality, multimodal planning, better management of existing systems, expanded public input and financial analysis requirements (Beimborn 1995).

Urban transport modelling is an important part of the transport planning process. Urban transport models meeting the new planning needs would contribute to the evaluation and design of policy options.

DESIRABLE CHARACTERISTICS OF AN URBAN TRANSPORT MODEL

A transport model is a simplified representation of a complex transport system. A good transport model should be based on sound economic theory, and be able to capture elements considered important for particular applications. Furthermore, the model should be transparent, with all assumptions clearly stated.

Models of special interest to this review are principally those of a strategic nature. That is, the type of model that can help determine corridor- or area-wide effects of policy options concerning investment, pricing and regulation of transport systems. In other words, such models are not intended for use in evaluating individual urban transport investment schemes. Evaluation of such schemes usually requires more detailed study, using, say, cost-benefit analysis, within the overall strategy suggested by the strategic model.

A number of characteristics are desirable from the strategic perspective. In particular, the model should:

- operate at a spatially aggregate level (zones). Models based on zonal aggregation provide a useful means of simplifying a complex transport system, although problems and processes associated with aggregation are not trivial. Simplification allows transport planners and policy-makers to concentrate on urban transport issues at a broader, more strategic level;
- have the capability of predicting long-term travel demand (20-30 years). Strategic planning requires a vision measured more in decades than in years. Capital inputs into transport infrastructure have long lives. Incorrect prediction of travel demand over the required period would waste scarce resources;
- be capable of explaining long-term changes in land use patterns. This would allow exploration of land use policy options as a way of addressing transport problems;
- permit close interaction between land use and transport because of their intricate relationship. Failure to do so would seriously undermine the accuracy and credibility of predictions made by the model; and
- be able to explain competition between various transport modes. The issue of mode choice has always been an important element in transport planning and policy making. Choice of mode has implications for the efficiency with which road space is used, as well as for the welfare of various road users. A good mode choice model (or sub-model) should be sensitive to those attributes that influence individual choice of transport mode.

HISTORICAL DEVELOPMENTS

A significant amount of research has been devoted to the study of urban transport systems over the past thirty years or so.

Early literature on transport modelling was dominated by the four-step, single-destination, separable-purpose and daily trip-based approach. In models such as these, the urban area is divided into a set of spatially contiguous trip-generating and trip-attracting zones (box 1.1). Travel demand is estimated using the sequential four-step process comprised of trip generation, trip distribution, modal split and route assignment. The system is characterised as being closed (due to lack of interaction with the land use system) as well as uni-directional because it does not allow any feedback from travel costs into the trip generation process.

The 1960s saw the emergence of behavioural demand models. Work in this area has accelerated since then, reflecting growing disenchantment with the conventional four-step approach. Behavioural demand models are based on economically rational notions of utility maximisation and consumer choice. They are also described as being 'disaggregate' because the unit of observation on which they operate is the individual traveller.

BOX 1.1 SPATIAL ZONES

In urban transport analysis, a study area is often divided into a set of contiguous trip-generating and trip-attracting zones. 'Trip-generating zones' (sometimes called trip origin or production zones) refer to those in which trips originate; 'trip-attracting zones' (also termed trip destination zones) refer to those where trips end. Representation of an urban area by a set of spatial zones simplifies empirical analysis of complex urban transport systems.

An urban area can be divided into a larger number of smaller zones, or a smaller number of larger zones. The ideal number of zones is usually decided empirically in specific situations. In general, the following factors are considered relevant in the design of a zoning system for a study area (Black 1981):

- Zones should contain distinctive land use patterns such as residential or industrial use;
- Characteristics of the activities within a zone should be as homogeneous as possible so that derived zonal means are representative of activity in the whole zone;
- The zone system should conform to census collection areas; and
- Zonal boundaries need to follow where possible, major roads, highways, rivers and other physical barriers to movement.

The choice of a zonal system is critical to the performance of models that utilise it. Different zoning systems may lead to different and sometimes contradictory conclusions. So, a general principle is to use as many zones as possible, in the hope that the internal homogeneity of the resultant zones and the differences between them can be maximised (Oppenheim 1995).

In earlier work, development of behavioural models concentrated on a single travel choice only, that of transport mode. Later research extended the modelling approach to cover other travel choices as well. Behavioural models provide complementary means (in the sense that they can be used to verify the validity of results generated from 'aggregate' analyses) as well as competing alternatives (when they are properly aggregated) to zone-based transport analysis.

Attempts were made in the 1970s to link transport models with land use models, leading to the development of so-called 'linked land use–transport models'. These models incorporate the effect of transport costs on the location decisions of households and firms. However, they fail to link, as in the case of the traditional four-step approach, total travel demand directly to travel costs.

Integrated land use–transport models began to gain popularity in the 1980s. A key feature of the integrated approach is that travel behaviour is modelled as a response to price signals, namely transport costs. The integration is achieved by explicit recognition of the two-way interaction between land use and transport systems, as well as by incorporating a wide range of theories (such as microeconomic, entropy or information, random utility, time geography, economic and welfare economics theories) and modelling techniques (such as spatial interaction, random utility and input-output models, and mathematical programming). Webster et al. (1988, pp. 31-37) provide a summary of the various theories and techniques. The integrated approach currently constitutes the state of 'best practice' in urban transport modelling and has been extensively used in transport planning and policy analysis, although scope exists for further improvement.

With the exception of behavioural travel demand models, which are based on individual data, the other three types of transport models rely on derived zonal data. While the main objective of this review is to examine zone-based models, a review of behavioural models is also necessary because they currently constitute the main competing alternative to aggregate zonal analysis. Moreover, many functional forms used in zonal analysis have their origin in behavioural demand models.

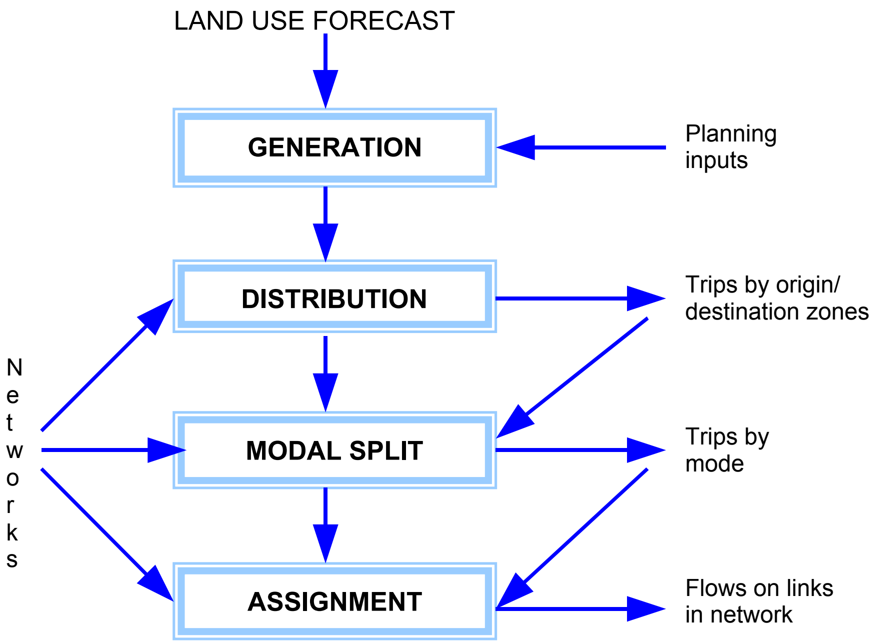
Finally, the activity-based approach focuses on space-time representation of daily or multi-day activity patterns of different types of households, and their implications for travel behaviour. Jones et al. (1983) provide coverage of the topic. An attempt by Pendyala et al. (1997) to apply microsimulation to the topic has brought a new perspective to activity-based analysis. However, the activity-based approach has yet to find its way into actual use within mainstream urban transport modelling (Southworth 1995). It is therefore not examined in this Working Paper.

CHAPTER 2 CONVENTIONAL FOUR-STEP TRANSPORT MODELS

The conventional four-step transport model is illustrated in figure 2.1. This approach provides a useful way to represent trip demand over complex networks with large numbers of competing destinations, modes and routes.

An important feature of the four-step model is that it is a recursive system with a uni-directional causal relationship. The sub-models first estimate the total number of trips generated, and then proceed to allocate them to destinations, transport modes and routes in the order shown in figure 2.1. The system is also meant to be iterative for the last three steps if travel costs are to vary under the congested situation. The iteration ensures that the predicted trip pattern is in equilibrium with travel costs on which it is based.

FIGURE 2.1 SCHEMATIC DIAGRAM OF A CONVENTIONAL FOUR-STEP URBAN TRANSPORT MODEL



Source Button 1977, p. 117.

TRIP GENERATION

Trip generation is the first of four sub-models of travel demand that are used in a conventional transport modelling process. The purpose of trip generation modelling is to determine the level of aggregate demand for trips originating in, and attracted to, each study zone.

It is usually assumed that the trip generation or attraction is determined solely by exogenous factors. The initial equations can therefore be written simply as:

$$T_i = f(S_i)$$

and

$$T_j = f(S_j) \tag{2.1}$$

where T_i and T_j represent the traffic generated by zone i and the traffic attracted to zone j respectively; and S_i and S_j are the socio-economic and/or land use characteristics of zones i and j.

Representation of zonal trip generation can be based on zonal means or totals. Zonal means should be used only when there is little variation in household or individual characteristics within the zone.

Two distinct techniques have been developed for estimating trip generation: multiple linear regression analysis, and category analysis or cross-classification analysis. Regression analysis relates travel demand (T_i or T_j) to variables such as those listed in table 2.1. An important drawback in the traditional trip generation process is the exclusion of the travel costs as an additional explanatory variable in the regression analysis. This has been due, in part, to the recursive nature of the system, a point analysed in more detail below.

Category analysis provides an alternative model of trip generation. It does not attempt to define explicit response surfaces as the regression analysis does, but rather is concerned with the construction of a multi-dimensional matrix, with each dimension representing an independent variable, stratified into a number of discrete classes or categories. Most empirical studies have assumed three socio-economic variables to have influence on household travel generation: these are car ownership; household size; and household incomes. There are a number of drawbacks to the category analysis method, including suppression of variances between household or zones in a specific cell, sensitivity of the results to grouping, and lack of an appropriate statistical measure to assess the reliability of the method (Stopher and Meyburg 1975).

TABLE 2.1 EXPLANATORY VARIABLES IN TRIP GENERATION MODELS

Socio-economic variables

- Car ownership
- Family size
- Number of persons five years old and over in the household
- Length of residence
- Family income
- Number of persons 16 years old and over
- Number of persons 16 years old and over who drive
- Age of head of household
- Distance from CBD
- Stage in the family life cycle
- Occupation of head of household
- Types of house structure

Land use variables

- Offices
- Industry
- Commerce
- Shops
- Education and health
- Public buildings
- Open space
- Transport and utilities
- Vacant land

Source Stopher and Meyburg (1975), pp. 111 and 121.

TRIP DISTRIBUTION

The main purpose of trip distribution modelling is to distribute the total number of trips originating in each zone among all possible destination zones available. As input, it uses a set of zonal trip productions and attractions, and attempts to estimate the way in which the production and attraction will be linked. The resulting trip distribution matrix can then be disaggregated by trip purpose (work trip, shopping trip, school trip, and so on) and time of day (peak and off-peak hours).

The trip distribution model may be expressed in a general form as follows:

$$T_{ij} = f(T_i, T_j, F_{ij}) \tag{2.2}$$

where T_{ij} is the traffic moving between zones i and j; and F_{ij} is the impedance to travel between i and j and can be represented by travel distance, time, costs, or a combination of them.

Gravity model

The most common form of trip-distribution model is the gravity model:

$$T_{ij} = \frac{kT_iT_j}{(F_{ij})^n} \tag{2.3}$$

where k, n = constants ($1<n<2$ usually)

Gravity models stipulate that the amount of traffic interaction between zone i and zone j is positively related to the product of the amount of traffic in zone i and zone j and inversely related to the impedance of getting from zone i to zone j. The k in equation 2.3 is a constant used to scale the estimate up or down, and the exponent n permits the friction or impedance factor to be manipulated in model estimation.

Equation 2.3 can also be used, through changes in trip attractors, as a travel demand forecasting tool in the case of two zones or cities (box 2.1).

The simple formulation of the gravity model in equation 2.3 suffers from a key weakness: the model cannot be consistently constrained to obey the trip-conservation rules described by equations 2.4 and 2.5.

Origin: $\sum_j T_{ij} = O_i$ (the number of trips originating in zone i) (2.4)

Destination: $\sum_i T_{ij} = D_j$ (the number of trips destined for zone j) (2.5)

Intuitively, equations 2.4 and 2.5 state that the sum of the trips over destinations j should equal the total number of trips originating from that zone and the sum of the trips over origins i should equal the number of trips attracted to that zone.

The problem can be overcome by replacing k in equation 2.3 with two sets of constants: one set associated with the production end of the trip and the other with the attraction end. This leads to a more respectable version of the gravity model that takes the form:

$$T_{ij} = O_iD_jA_iB_jf(c_{ij}) \tag{2.6}$$

BOX 2.1 APPLICATION OF THE GRAVITY MODEL

Gravity models are based on the Newtonian concept. Bodies of large mass (large cities) have more attraction for people than those of less mass (small towns). So more people will visit and move to large places than to small ones. The Newtonian concept of mass can be represented in terms of population, total household incomes, level of economic activity, and so forth. Distance is an impedance factor preventing two bodies from interacting. In modern transport analysis, the distance variable is usually replaced by the generalised costs of transport (value of time, cost of petrol, etc).
BTCE (1997) and BTE (1998) used a variant of the standard gravity model (2.3) to estimate demand for passenger travel on 10 pairs of major interregional links in Australia. The model has the form:

$$Passenger\ Travel_{o-d} = \frac{(Population_o * population_d)^b}{(Travel\quad costs / weekly\quad earnings)^a}$$

where the impedance factor is represented by the ratio of generalised travel costs over weekly earnings, or the number of weeks of earnings required to pay for the trip. The model was estimated using the panel data between 1970-71 and 1995-96. A set of dummy variables was used to account for differences in levels of travel demand between the ten links and special events. *b* was estimated as 0.5, and the exponent of the impedance variable *a* was estimated to be 1.25. The model successfully explains about 97 per cent of the variation in total passenger travel demand over the ten corridors.

where $f(c_{ij})$ represents some generalised cost function of travelling between i and j; A_i and B_j are constants associated with the production and attraction zones respectively and are defined as follows:

$$A_i = [\sum_j D_j B_j f(c_{ij})]^{-1} \quad \text{and} \quad B_j = [\sum_i O_i A_i f(c_{ij})]^{-1} \qquad (2.7)$$

Note that terms A_i and B_j in equation 2.7 are mutually dependent and need to be solved iteratively. The process can be implemented by first making all B_j equal to 1 and calculate the values of A_i. The calculated values of A_i are then inserted into the equation of B_j to obtain new values of B_j, with the process being repeated until numerical equilibrium is reached.

Also note that the crude notion of impedance being a simple function of distance has been replaced by the notion of 'generalised costs' (appendix I), which includes both travel time and any fares or other monetary operating costs incurred during the trip. In doing so, a new constraint must be met:

$$\sum_{i}\sum_{j} T_{ij}c_{ij} = C \tag{2.8}$$

Equation 2.8 states that there is a total cost of transport, C, which must exactly equal the sum of all trips multiplied by the cost of transport for each origin-destination pair c_{ij}.

A number of specifications for the cost function are possible, but the most common ones used in transport analysis are:

Exponential function: $f(c_{ij}) = \exp(-\beta c_{ij})$ (2.9)

Tanner's function: $f(c_{ij}) = c_{ij}^{\alpha}\exp(-\beta c_{ij})$ (2.10)

Equations 2.9 and 2.10 both belong to the family of distance-based cost decay functions with one or more parameters for calibration.

Entropy-maximising method

In a major theoretical contribution, Wilson (1970) demonstrated how the gravity model could be derived from entropy-maximising methods. The entropy concept for generating models of spatial interaction involves the notion of uncertainty in a system (of 'molecules') subject to random motion. The most probable state of the elements of the system is said to be achieved when its entropy is maximised subject to the constraints on the micro-state of the system (box 2.2). The entropy method, when applied to urban transport studies, is concerned with first determining an expression to calculate the number of possible ways that can give rise to a set of trip matrices and then finding a trip matrix with the maximum probability of occurrence subject to origin and destination constraints (box 2.3). This maximisation process results in a formulation similar to the modified version of the gravity model described above.

Equation 2.6 is called the 'doubly constrained model' in Wilson's entropy maximising process. It has been popular for modelling journeys to work. Wilson also derived three other cases of the entropy-maximising spatial interaction model, depending on the availability of information.

Origin (production) constrained

In the origin constraint case only the origins of the flows are known, not their destinations. The destination term D_j in equation 2.6 must be replaced by a hypothetical indicator of the attractiveness of zone j. Origin-constrained models contain a term for each production zone A_i and this zonal variable ensures that when the trips in each row of the estimated trip matrix are added up, the estimates of zonal production equal the actual zonal traffic production. However, the column totals for each zone — the estimates of zonal traffic attraction — do not equal the actual zonal traffic attraction. The origin-constrained interaction model can be used to represent choices of trip destinations for particular purposes (such as work, shopping, recreation and so on).

Destination (attraction) constrained

In the destination constraint case, only the destinations of the flows are known, but not their origins. The origin term O_i in equation 2.6 must be replaced by a hypothetical indicator of the attractiveness of zone i. Destination-constrained models contain a term for each attraction zone B_j, and this zonal variable ensures that when the trips in each column of the estimated trip matrix are added up, the estimates of zonal attraction equal the actual zonal traffic attraction. However, the row totals for each zone — the estimates of zonal traffic production — do not equal the actual zonal traffic production. The destination-constrained interaction model can be used to represent choices of origins for particular purposes (such as residence).

Unconstrained

The unconstrained case corresponds to the situation of minimum information where neither origins nor destinations are known. It is assumed that values of attractiveness need to be given to both origin and destination zones. The specification can be reduced to the simple gravity model with a single k. This constant term is adjusted to ensure that the estimate for the total number of trips in the O-D matrix equals the total number of trips in the actual O-D matrix.

The entropy method is probabilistic in nature (box 2.3). It provides an alternative explanation for the gravity model. However, being an analogy, the approach does not contribute much in a conceptual or behavioural sense. As will be shown in the next chapter, there are more conceptually satisfactory approaches that yield identical model specifications.

MODAL SPLIT

'Modelling modal split' refers to the allocation or 'distribution' of trips between the various modes available. The information available for building mode-choice models is the observed modal split, characteristics of the travelling population and the operational characteristics of the competing transport modes.

The general form of the modal split model can be written as:

BOX 2.2 WHAT IS ENTROPY MAXIMISATION?

The term 'entropy' comes from thermodynamics where it provides a quantitative basis for the common observation that naturally occurring processes have a particular direction. For example, if a vessel containing a gas were connected by a pipe to an evacuated vessel, the gas would expand to fill all the available space uniformly.

Once the atomic theory of matter is accepted, the laws of thermodynamics are explained in terms of the statistical behaviour of large collections of molecules, using the theory of statistical mechanics. It is impossible to know the exact state of all the molecules; it is only possible to observe the gas in bulk. The exact state of the molecules is called the 'microstate'. The bulk state is called the 'macrostate'. Each microstate has bulk properties corresponding to one of the macrostates. When the gas is in a particular macrostate, it also corresponds to one of the microstates, but there is no practical way of knowing which. Assuming that each microstate is equally probable, the probability of a given macrostate is proportional to the number of microstates.

Entropy can be shown to be a measure of the number of microstates that a system can assume. Imagine a case where there are four gas molecules in a bottle which are pushed into another bottle with two compartments. There are five possible ways in which the molecules can be divided up between the two compartments as shown in the table below. These are the macrostates. The first macrostate, with all four molecules in compartment one, can only occur in one way. However, for the second, with three molecules in compartment one and one molecule in compartment two, there are four microstates. The full list of possibilities is shown in the table below.

	Macrostates				
Compartment 1	4	3	2	1	0
Compartment 2	0	1	2	3	4
No. of microstates	1	4	6	4	1
Probability	0.0625	0.25	0.375	0.25	0.0625

The total number of possible microstates is 16. Assuming that each microstate is equally probable, the probability that each macrostate will occur is on the bottom line of the table. The macrostate with two molecules in each compartment is the most probable final state. A system will tend to equilibrium in its most probable final state, that is, where entropy is at a maximum.

As the number of molecules increases, the macrostate having all molecules in one compartment can still only occur in one way. However, the number of ways the molecules can be distributed between the compartments increases rapidly. In normal situations the number of molecules is so great that the maximum entropy macrostate is much more probable than others.

Entropy is fundamentally a statistical concept. This has allowed it to find applications outside physics, such as in information theory, and transport analysis.

$$T_{ijm} = f(I_{ij1},...,I_{ijm},T_{ij}) \tag{2.11}$$

where T_{ijm} is the traffic from i to j using mode m. I_{ijm} represents the attributes of various competing modes.

There are basically two types of modal-split model: 'trip-end modal split models' and 'trip-interchange modal split models'.

Trip-end models perform modal split estimation immediately after the trip generation process has been completed. They split total travel demand for each zone by transport mode. Most of the trip-end modal split models use either regression analysis such as that used in the trip generation process, or analytical models similar to the gravity model for trip distribution. In earlier research on this topic, analyses based on trip-end models concentrated entirely on social and economic characteristics of the study areas, and the relative attractiveness (costs) of the various modes was excluded from the list of explanatory variables.

The trip-interchange model, which operates at the stage between trip distribution and route assignment, has as its function the splitting of specific intra- and inter-zonal trips among available modes. The key reason for the development of trip-interchange models is the relative unresponsiveness of the trip-end models to the transit system. Therefore, trip-interchange models tend to incorporate the comparative time, cost and service differentials between competing modes into the regression analysis.

There are many techniques available for modelling modal choice, but the most appealing one seems to be logit (either binomial or multinomial) analysis which is based on random utility models, and is consistent with the theory of consumer behaviour. This approach is detailed in Chapter 3.

TRAFFIC ASSIGNMENT

Traffic assignment modelling distributes traffic among the routes of an urban transport network. Separate assignments are made for each of the different travel modes. Mathematically, this step can be expressed in general form as follows:

$$T_{ijmp} = f(I_{ijm1},...,I_{ijmp},T_{ijm}) \tag{2.12}$$

where T_{ijmp} is traffic using route p when travelling by mode m between i and j; and $I_{ijm1...p}$ represent the characteristics of various paths.

Earlier approaches focused on single-path assignment, assuming an infinite capacity for each network link (or 'zero-flow' network travel times). The method is to first identify the least-cost route from the given origin and

BOX 2.3 THE WILSON MODEL

The Wilson (1970) model is based on the concept of entropy which has its origin in thermodynamics. Cesario (1975) provides a simple illustration of how the entropy method is applied to transport analysis.

Consider the case of the 2x2 home-to-work origin-destination table shown below.

Destinations

		Zone 1	Zone 2	O_i
Origins	Zone 1	t_{11}	t_{12}	3
	Zone 2	t_{21}	t_{22}	3
	D_j	4	2	$T = 6$

In the table, the number of trips originating in zone i (O_i) is known, as well as the number of trips ending in zone j (D_j). The total number of trips (T) is 6. The problem is to find a distribution (t_{11}, t_{12}, t_{21}, t_{22}) which is most likely to occur, given the amount of information available.

There are only three possible trip matrices that satisfy the origin and destination constraints described in equations 2.4 and 2.5.

$$T = \begin{matrix} 3 & 0 \\ 1 & 2 \end{matrix} \qquad T = \begin{matrix} 2 & 1 \\ 2 & 1 \end{matrix} \qquad T = \begin{matrix} 1 & 2 \\ 3 & 0 \end{matrix}$$

(a) (b) (c)

These three possible outcomes are called 'macrostates'. A macrostate specifies how many people travel between i and j (that is, t_{ij}) without considering 'who travels where'. Associated with each of these macrostates is a series of 'microstates', the sum of which forms a complete specification of the system.

In order to determine which outcome is most likely, there is a need to calculate the total number of microstates associated with each of the three macrostates. This can be achieved by using the familiar combinatorial formula of statistics (Fraser 1958) shown below:

$$W = \frac{T!}{t_{11}!t_{12}!t_{21}!t_{22}!} = \frac{T!}{\prod_{ij} t_{ij}}$$

where ! represents factorial operation. Using this equation, W is found to be 60, 180 and 60 for (a), (b) and (c) respectively. Assuming that each microstate is equally probable under entropy, the most probable macrostate is given by (b).

The Wilson model is based on a similar concept, except that the function to be maximised adopts a slightly different form and there is the additional cost constraint shown in equation 2.8.

destination pairs, and then to assign the total number of trips to that route on an 'all-or-nothing' basis. This method wins merit for its theoretical underpinning but fails on its assumptions such as 'zero-flow' on road and perfect information among motorists. The omission of capacity constraints tends to lead to peculiar and unrealistic results, because traffic will be assigned only to those links that lie on the minimum cost path, thereby overloading these links.

The introduction of capacity constraints leads to multiple-path assignment models that conform with the first of Wardrop's (1952) two principles (box 2.4). The basic idea of these models is, that in a congested situation, traffic would be spread over all the alternative transport routes between a particular zonal pair in such a way that travel time or costs become equal for using these different routes. The result is an equilibrium that is optimal from the viewpoint of each user. No individual traveller can improve his or her utility by finding a route with less travel time or lower travel costs.

An alternative solution to the trip assignment problem is to use a mechanism to obtain the probabilities of route choice. In general, this approach does not allocate traffic to the minimum path only, but to the *n*-shortest paths instead. This can be achieved by using a multinomial logit model with generalised cost included as a key explanatory variable. Under this approach, the higher the cost to travel on a particular route, the less the probability of that route being chosen for travel. The probabilistic assignment approach has received considerable attention in recent years, especially where there has been a tendency to model travel demand in a consistent manner, such as under the framework of random-utility-based travel demand models.

CRITIQUE OF CONVENTIONAL TRANSPORT MODELS

Four-step transport models are conventionally applied in a uni-directional manner, commencing with trip generation and ending with traffic assignment. This can clearly be seen if the four sub-models are restated in their most basic form as equations 2.13-2.16.

Generation	$T_{i-or-j} = f(S_{i-or-j})$	(2.13)
Distribution	$T_{ij} = f(T_i, T_j, ...)$	(2.14)
Modal split	$T_{ijm} = f(T_{ij}, ...)$	(2.15)
Assignment	$T_{ijmp} = f(T_{ijm}, ...)$	(2.16)

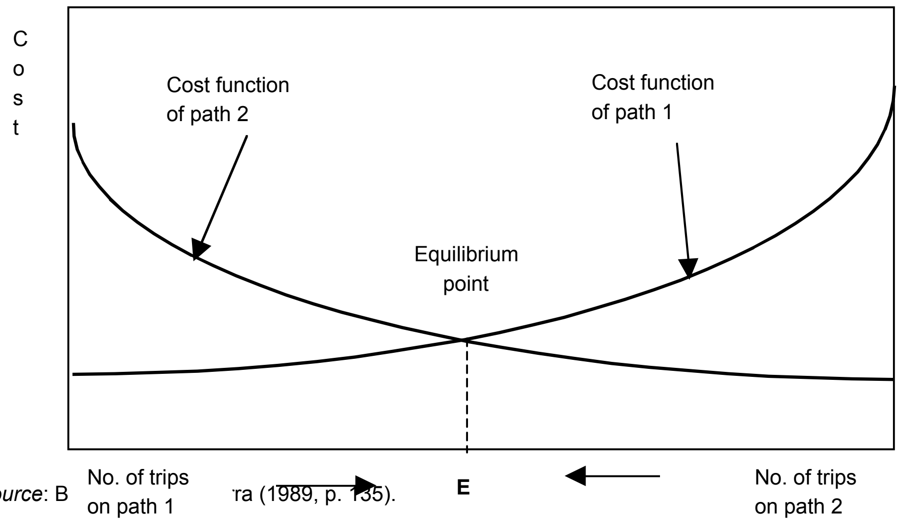
Such a system assumes that each process is independent of the process below it. This assumption leads to the exclusion of the cost of travel from the list of determinants affecting the level of travel demand. Consequently, forecasts made by the trip generation model are based only upon changes in income,

BOX 2.4 WARDROP'S PRINCIPLES

Wardrop (1952) introduced two principles of equilibrium for transport systems. They are summarised neatly by Friesz and Harker (1985, p. 162) as follows:

> **Wardrop's first principle** Each user non-cooperatively seeks to minimise his cost of transportation. A network flow pattern consistent with this principle is called a 'user-optimised equilibrium'. Specifically, a user-optimised equilibrium is reached when no user may lower his transportation cost through unilateral action.
>
> **Wardrop's second principle** The total cost of transportation in the system is minimised. A network flow pattern consistent with this principle is called a 'system-optimised equilibrium' and requires that users cooperate fully or that a central authority controls the transportation system. Specifically, a system-optimised equilibrium is reached when the marginal total costs of transportation alternatives are equal.

The equilibrium assignment method is an application of Wardrop's first principle. It can be illustrated with the help of the figure below. For simplicity assume that there are two possible paths connecting an origin *i* and destination *j*. The total number of trips is represented by the horizontal axis. The two cost functions (one is associated with path 1 and the other with path 2) have the usual upward-sloping shape reflecting the growing cost as the result of a larger number of users joining these paths. The two cost curves intersect at point *E* where the perceived cost of the two paths is equal. At *E*, no user can improve his or her utility or lower his or her cost by switching to another route. At any other point different from *E*, a certain number of users would benefit by switching from one path to the other. Cost minimisation behaviour by users ensures that equilibrium will always be reached.



Source: B ra (1989, p. 135).

taste and other variables that cause shifts in the demand curve, excluding factors that cause movements along it (Button 1977, p. 123). The failure to generate parameters that specify the price-responsiveness of travel demand makes the traditional approach impotent in any price-based policy analysis.

Another drawback of conventional transport models is that they lack a consistent rationale that would jointly address all aspects of demand (Oppenheim 1995). Travel demand at the four different stages is analysed by using different approaches. Trip generation is typically performed by using linear regression based on *ad hoc* specifications; trip distribution is largely based on spatial interaction models that rely on gravity formulas; modal split and route assignment are generally given to more behavioural interpretation, but they are not integrated. There is a strong argument for integrating the various approaches, especially when it has been shown that there is a close relationship between the gravity model and behavioural (ie. logit) models, both functionally and numerically (Anas 1983a).

The third drawback, which is related to the second, is the lack of a consistent measure of welfare. Consumer surplus may be derived, say, at the modal split level by using a behavioural demand model, but not for other stages, due to the different analytical frameworks used.

A final drawback is that the analysis is based on statistically derived zonal correlations. Use of grouped data tends to underestimate variances of variables and may therefore lead to 'fallacies of inferences'. While this shortcoming applies to any zone-based analysis, the types of aggregation used in the disaggregate modelling (discussed in the next chapter) help to minimise the risk of incorrect inferences.

However, not all four-step models are of the traditional, conventional variety. Some incorporate considerable sophistication. Care needs to be taken, therefore, to judge individual models on their merits.

CHAPTER 3 BEHAVIOURAL TRAVEL-DEMAND MODELS

Criticism of conventional transport models spurred intense research from the 1970s onwards into behavioural demand models. The breakthrough was made by McFadden (1973) and Ben-Akiva (1973) who formulated the problems of travel mode and location decisions as problems in micro-economic consumer choice among discrete alternatives.

A behavioural model involves the representation of individual choice when transport users are confronted with alternatives. Travel-related choices include trip frequency, time of travel, destination, transport mode and routes.

Behavioural travel-demand models differ from conventional four-step transport models in two important ways. Analysis is carried out at a decision-making level, such as a person or household, rather than on the derived zonal traffic. Models are thus derived from the micro-economic theory of consumer behaviour, rather than on the basis of *ad hoc* specifications.

Up to the mid-1980s, it was accepted almost unequivocally that modelling demand for travel should be based on information about observed choices, ie, revealed preference data (Ortuzar and Willumsen 1994). Since the end of the 1970s, however, a complementary development of the stated preference approach has occurred, aimed at overcoming shortcomings of the revealed preferences approach, such as the inability to embrace new options in the travel choice set. Hensher (1997) provides an excellent exposition of the stated preference approach in transport analysis. The current review focuses on models that are based on revealed preference data.

RANDOM UTILITY AND MULTINOMIAL LOGIT MODELS

Travel choice models are based on the concept of random utility. The theory says that the utility derived by consumer *i* who chooses option *k* contains both a systematic component $V_k^i(X_k, S_i)$ (which is the average utility of alternative *k* for individual *i* with socioeconomic characteristics S_i), and an unobservable component $\varepsilon(X_k^{'}, S_i)$ (which is the individual utility of alternative *k* for individual *i* with socioeconomic characteristics S_i). The random utility function is expressed in the following equation.

$$U_k^i = V_k^i(X_k, S_i) + \varepsilon(X_k^{'}, S_i) \qquad (3.1)$$

where X_k and $X_k^{'}$ contain observable and unobservable attributes respectively of alternative *k*.

Consider an example of binomial choice of transport modes. Suppose that a commuter has to choose between 'driving to work' and 'taking public transport', and that the systematic portions of utility from the two alternatives are:

$$V_1^i(X_1, S_i) = -10 \qquad \text{(alternative 1: driving to work)}$$

$$V_2^i(X_2, S_i) = -15 \qquad \text{(alternative 2: taking public transport to work)}$$

The utilities are negative, indicating that both driving and taking public transport to work produce disutility. If utility were not random, then it would not be difficult to predict that commuter *i* would choose to drive to work, because driving yields less disutility. However, if random components are introduced in the utility function, as in the case of equation 3.1, the total utilities of alternatives become unknown, thereby requiring alternative methods to determine the choice that would be made.

If the consumer makes choice *k* in particular in equation 3.1, we assume that U_k^i is the maximum among the *j* utilities where *j*=1, …, *J*. Hence the statistical model is driven by the probability that choice *k* is made. This probability is given as:

$$\begin{aligned} P_k^i &= \text{Prob}[U_k^i \ ? \ U_j^i] \\ &= \text{Prob}[V_k^i + \varepsilon_k^i \ ? \ V_j^i + \varepsilon_j^i] \\ &= \text{Prob}[\varepsilon_j^i - \varepsilon_k^i \leq V_k^i - V_j^i] \qquad \text{For all } j ? \ k \end{aligned} \qquad (3.2)$$

The model is made operational by a particular choice of distribution for the random component of the utility function. The simplest means of obtaining a solution to equation 3.2 is to assume a Weibull distribution (also called the 'double-exponential' or 'extreme value type I' distribution) for disturbances $\varepsilon(X_k^{'}, S_i)$. Other assumptions can be made about the random component of the utility function, one of which is that the disturbances are normally distributed, leading to probit models. Probit models are claimed to be able to produce more realistic results (Daganzo 1980), but they face considerable operational difficulties compared with logit models.

Using the Weibull function leads to McFadden's logit model, which takes the form:

$$P_k^i = \frac{\exp[V_k^i(X_k, S_i)]}{\underset{j}{}\exp[V_j^i(X_j, S_i)]} \qquad (3.3)$$

where P_k^i is the probability that individual *i* will choose alternative *k*; *X* and *S* are two vectors representing choice-specific attributes and individual-specific characteristics respectively. Choice-specific attributes could include, for example at the modal choice level, costs associated with different transport modes, time required for the journey, comfort, and so on. Individual-specific characteristics could include, among other things, level of income, car ownership and household size.

In essence, equation 3.3 basically states that the probability of an individual choosing an alternative k from a set of available alternatives is a function of the attributes of the available alternatives and his or her characteristics.

However, when the data consist of choice-specific attributes only, equation 3.3 can be reduced to equation 3.4 which is the conditional logit model (Greene 1991). It is assumed in this model that the characteristics of the individual enter the utility function through the vector of coefficients of X_j's.

$$P_k^i = \frac{\exp[V_k^i(X_k)]}{\sum_j \exp[V_j^i(X_j)]} \qquad (3.4)$$

In general, X_j's in V_j^i are assumed to be linearly additive and hence equation 3.4 can be written as:

$$P_k^i = \frac{\exp(\beta' X_k)}{\sum_j \exp(\beta' X_j)} \qquad (3.5)$$

Intuitively, equation 3.5 says the probability that the individual i will choose the alternative k is a function of overall desirability (or attributes) of the chosen alternative, relative to all other alternatives j. The relationship is regulated by the vector of coefficients, β, the role of which is to set the 'dispersion' of the distribution of the utility function. Higher values of β mean individuals are more sensitive to travel costs or are able to capitalise more on differences in attributes among alternatives. As the values of β approach infinity, the probability, P_k^i, will tend to one. On the other hand, when $\beta=0$ (that is, individuals are not sensitive to costs or choice attributes), the probabilities of all options become equal. That is, $P_k^i = 1/J$, where J is the total number of options, so that P_k^i is uniformly distributed.

Equation 3.5 has the restrictive property of 'independence from irrelevant alternatives' (IIA). That is, the ratio of the probabilities of an individual selecting two alternatives is independent of the remaining probabilities. To see this, consider the 'red-bus-blue-bus' conundrum (Mayberry 1970). Assume that a commuter has to choose between 'driving a car' and 'taking a red bus', and further suppose that the ratio of the probability this commuter selects 'driving a car' (0.3) to 'taking a red bus' (0.1) is three to one (that is, a commuter is three times as likely to select 'driving a car' as to 'taking a red bus'). Now assume that a new bus option is introduced, such as a blue bus. Because of differential competition effects, it would seem logical to expect the 10 per cent of the commuters who take a (red) bus to work to distribute evenly between the red and blue buses, and the others to continue as usual. However, if they do, the ratio of the probabilities of 'driving a car' over 'taking a red or blue bus' will increase to 6 to 1. In order to preserve the 3 to 1 odds ratio between car and buses, commuters travelling by car and the red bus need to switch to the blue bus. The IIA property has been perceived as a disadvantage which makes the multinomial logit model fail in the presence of correlated alternatives (Ortuzar and Willumsen 1994).

It should be noted that the random utility model expressed in the logit form (equation 3.5) is almost identical to the entropy maximising model (equation 2.6 with the cost function specified as equation 2.9) described in Chapter 2. The inextricable relationship between the two types of models can be considered as a result of similar assumptions (but in different order) being made in their derivations. In entropy models, choices are first assumed to be perfectly random, then a rational (cost) restriction is introduced. In contrast, the random utility model begins by assuming that choices are perfectly rational, then, because of aggregation, introduces random elements (de la Barra 1989). Behavioural models since their inception, have been found to be more acceptable, as they possess some desirable properties that entropy interaction models don't have.

ELASTICITY PROPERTIES OF MULTINOMIAL LOGIT MODELS

The logit model has some highly acceptable elasticity properties. Given equation 3.5, the direct elasticity of probability for alternative k with respect to characteristics x_{ks} of that alternative can be shown to have the form (Stopher and Meyburg 1975, p. 283):

$$\eta^i_{kks} = b^i_s x_{ks}(1 - P^i_k) \qquad (3.6)$$

The intuition of equation 3.6 is as follows: the first two terms tell us that the elasticity is a function of the amount of a particular attribute possessed by alternative k and its weight b^i_s in the utility function. The last term modifies the strength of the direct elasticity by the market share alternative k has not yet obtained $(1 - P^i_k)$. An example of direct elasticity of probability is given in table one (box 3.1), where all diagonal elements correspond to the concept of a conventional own-price elasticity of demand.

Similarly, the cross-elasticity of probability may be expressed as a function of the model parameters. Thus, the cross-elasticity of demand for alternative k with respect to attribute s of alternative m is given as follows:

$$\eta^i_{kms} = -b^i_s x_{ms} P^i_m \qquad (3.7)$$

Intuitively, the cross-elasticity of demand for alternative k with respect to alternative m with attribute s, is defined as being the product of the amount of the attribute possessed by alternative m, the weight of that attribute in negative form and the market share that alternative m has. Application of cross-elasticity values can also be found in table two (box 3.1) where all off-diagonal elements show the effects of changes in other prices on the probability of a particular mode being chosen.

The concept of elasticity is important. If attributes of each alternative are represented by generalised costs, the estimates derived on the basis of equations 3.6 and 3.7 will be similar to the concept of own- and cross-price elasticity of demand used in conventional micro-economic theory (box 3.1). These estimates are of considerable interest to transport planners and policy makers.

BOX 3.1 APPLICATION OF THE MULTINOMIAL LOGIT MODEL IN MODE CHOICE

Hensher (1986) used a model based on equation 3.5 to analyse mode choice behaviour for urban travel for a sample of Sydney commuters. The four choices were defined as: car driver (CD), car passenger (CP), train and bus. The attributes of the four alternatives include, among other things, in-vehicle costs, in-vehicle time, walk time, waiting time and parking costs. The sample consists of 1455 observations.

The four predicted probabilities and frequencies are estimated and given in the table below: 'car driver' dominated modal shares.

Predicted probabilities and frequencies

	CD	CP	Train	Bus
Probability	0.8863	0.0380	0.0139	0.0618
Predicted frequency	1,290	55	20	90
Actual frequency	953	78	279	145

Source Hensher (1986).

Hensher also calculated the elasticities of probabilities, one being with respect to in-vehicle costs (see table below). As expected, the direct elasticities (diagonal elements) are all negative, implying that demand for a particular mode is inversely related to the costs associated with that mode. The positive numbers for cross-elasticities (off-diagonal elements) indicate that the four choices are substitutes.

Elasticities of probabilities with respect to in-vehicle costs

	CD	CP	Train	Bus
CD	-0.077	0.253	0.253	0.253
CP	0.002	-0.013	0.002	0.002
Train	0.098	0.098	-0.231	0.098
Bus	0.042	0.042	0.042	-0.292

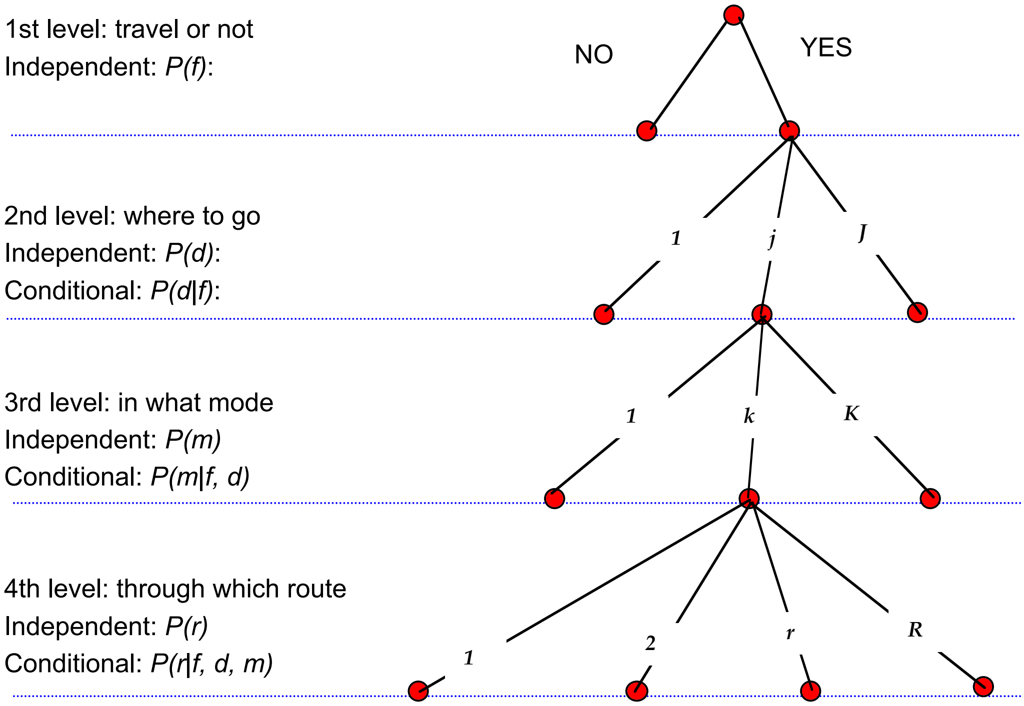
Source Hensher (1986).

HIERARCHICAL LOGIT MODELS

The foregoing discussion relates only to a single-level choice. For instance, the choice of mode given the destination, or the choice of route given the mode. In an urban transport system, however, travellers are likely to face a number of choices that are related to each other. This gives rise to the issue of the structure of travellers' choice processes.

The issue can be illustrated with the help of a hierarchical choice decision tree. Adopting the standard sequential assumptions in the conventional four-step approach leads to a decision tree that looks like that shown in figure 3.1. The process is concerned with estimating the probability that an individual will undertake a trip with frequency *f* to destination *d* using mode *m* via route *r*.

FIGURE 3.1 THE STRUCTURE OF THE TRAVELLER'S CHOICE PROCESS



Source Based on Oppenheim (1995), p. 28.

Stopher and Meyburg (1975, pp. 284-288) suggested three types of model structure with respect to choice models. One type of structure is an independent structure. It requires the assumptions of independence and separability, both being implied by the axiom of IIA. The independence assumption implies that there is no dependence of one choice on another. That is, the choice of mode has no impact on the choice of route, or of destination. The separability assumption implies that there is no correlation among four sets of attributes, X_f, X_d, X_m and X_r. Under the assumptions of total independence and separability, the probability for each link of the decision chain can be estimated separately through independent multinomial logit models ($P(f)$, $P(d)$, $P(m)$, $P(r)$); and if the joint probability (that is, say, the probability of making a trip to a particular destination by a certain mode through a specific route) is of interest, it is simply a matter of multiplying the probabilities at each stage in the hierarchy (*P(f)*P(d)*P(m)*P(r)*). It should be noted that, while independence and separability assumptions offer analytical and computational convenience, they prove in practice to be too restrictive. In the case of transport demand, decisions as to whether or not to make a trip, where to go, in what mode and through which route, are likely to be made simultaneously.

Another type of modelling structure is a simultaneous one. It assumes that decisions are not multiple but single. The simultaneous approach requires, for the evaluation of a specific probability, the definition and estimation of all alternative travel choices based on all relevant characteristics, that is, X_{fdmr}. Evidently, this could lead to a very large and complex model in which there are likely to be serious estimation problems.

The third type of modelling structure is a recursive and sequential one, leading to hierarchical or nested logit models. This approach recognises in some way the simultaneous nature of the process, but prunes less relevant alternatives within and between travel choices, on the assumption that the conditional probability for a given choice depends only on a part of the total utility function. The result is a set of conditional probabilities such as those shown in figure 3.1. The joint probability of a specific trip being made is given by the product of *conditional* probabilities, as shown in equation 3.8.

$$P(f, d, m, r) = P(f)*P(d|f)*P(m|f, d)*P(r|f, d, m) \qquad (3.8)$$

In this recursive structure, the estimation process must begin with the last link of the decision chain (route choice, in this instance), then proceed back to the top link (frequency) in order to ensure that the strict utilities are preserved throughout the process. An example illustrates the procedure.

Consider a case of bi-dimensional choices, such as the combination of destination (*d*) and mode (*m*) choice. In such a context, the utility of the destination level *d* and mode level *m* joint choice can adopt the following form (Ortuzar and Willumsen 1994; Williams and Ortuzar 1982):

$$U_{dm} = V_d + V_{dm} + \varepsilon_d + \varepsilon_{dm} \qquad (3.9)$$

where the component V_d represents the portion of the utility specific to the destination choice *d* and the component V_{dm}, to the disutility associated with the cost of travelling using mode *m* given the destination. The random utility terms ε_d and ε_{dm} have similar interpretations.

It can be shown that if the random terms ε are separately IID (independent and identically distributed), under certain conditions, the hierarchical logit model (Williams 1977; Daly and Zachary 1978) is formed:

$$P(d,m) = \frac{\exp\{\beta(V_d + W_d)\}}{{}_{d}\ \exp\{\beta(V_d + W_d)\}} * \frac{\exp(\lambda V_{dm})}{{}_{m}\ \exp(\lambda V_{dm})} \qquad (3.10)$$

where

$$W_d = \frac{1}{\lambda} \ln \;{}_{m}\; \exp(\lambda V_{dm}) \qquad (3.11)$$

and λ has an interpretation similar to β in the destination choice model.

In estimating equation 3.10, the first step is to estimate the second component on the right hand-side of the equation. That is, the mode choice. This gives rise to the expected maximum utility of the mode nest options (equation 3.11). This maximum utility is called the 'inclusive value'. The next step is to estimate the destination choice model including the inclusive value variable obtained from the first step. The final step is to estimate the joint probability by multiplying the probabilities for the destination and the mode.

Hierarchical logit models provide a very useful way to overcome the problems of attribute correlation associated with the multinomial logit models. In the example of 'red-bus-blue-bus' conundrum, if a hierarchical approach is adopted — that is, the first level choice is made between bus and car and the second level choice is made between red bus and blue bus, the wrong inference can certainly be avoided. However, in a practical sense, making decisions on nesting patterns is by no means an easy task, especially when a large number of spatial choices need to be made (Fotheringham and O'Kelly 1989). It is important that the nesting patterns be appropriate, as they have implications for the results obtained from the model.

Hensher (1993) and BTCE (1996, appendix V) provide a readable description of the design of the hierarchical logit model used in BTCE (1996) to represent urban passenger behaviour.

EXPECTED MAXIMUM UTILITY AND CONSUMER SURPLUS

Given equation 3.5, the expected maximum utility of the choice set can be calculated at the various decision levels shown in figure 3.1 according to the formula:

$$W = \frac{1}{\beta} \ln \;{}_{j}\; e^{\beta' X_j} \qquad (3.12)$$

where W is the expected maximum utility of a given choice set; β is the set of parameters for the exponential function; and X_j is the vector of choice-specific attributes. If X_j are represented by travel costs alone, then equation 3.12 becomes a function for estimating the expected minimum disutility or composite costs (de la Barra 1989).

The expected maximum utility is also known as the inclusive value. The inclusive value serves as a summary measure of the desirability of the entire choice set. Under a hierarchical structure, inclusive values must be calculated by starting at the bottom of the tree (the 'route' level in figure 3.1) and working upwards. Once found, these inclusive values are used in the estimation of probabilities, a process which is performed in the reverse order, from the top downwards. Inclusive values can be aggregated from the bottom to the top representing the level of utility or user benefit at the different choice levels.

Inclusive values are related directly to a conventional measure of welfare: consumer surplus. Changes in consumer surplus resulting from an exogenous change can be derived from the random utility model, as the difference between the aggregate of the inclusive values before and after the policy change. This change is given as:

$$\Delta W = \frac{-1}{\beta} \ln \frac{\sum_j e^{\beta X_j^i} (after \quad policy \quad change)}{\sum_j e^{\beta X_j^i} (before \quad policy \quad change)} \tag{3.13}$$

The indicator ΔW is conceptually equivalent to the traditional consumer's surplus indicator (Small and Rosen 1981), and has been widely used in policy evaluation such as assessment of impacts of schemes based on congestion pricing.

AGGREGATION

Because of the number of desirable properties that the random utility model possesses, attempts have been made to improve methods for predicting aggregate travel demand based on disaggregate behavioural relationships (Koppelman 1984; Daly and Ortuzar 1990). Koppelman (1984, pp. 19-60) examines a number of procedures for aggregation with different information requirements and computational complexity, and discusses their implications for aggregation error. These include naive, classification, enumeration, integration and statistical differential procedures. The latter two approaches were developed mainly to cater for the need to aggregate non-linear behaviour in travel demand models.

Naive

The 'naive' procedure uses the representative (average) values of explanatory variables in the disaggregate probability model to obtain the aggregate share of choosing an alternative. While this approach requires minimum data and the least computational effort, it could introduce errors in prediction if variable values are widely dispersed around their mean, or if the relationship is non-linear.

Classification

One way to reduce variances of explanatory variables is to have the population classified into relatively homogeneous groups, the average values of variables of which are then used to represent aggregate demand for each group. The group aggregate demands are finally aggregated using appropriate weights for each group. This is called the classification procedure. The classification procedure requires more data than the naive approach, because it requires information about the expected size of each class, as well as the average value of the variable for each class. Increasing the number of classes leads to an improvement in aggregate prediction, but imposes an increase in data requirements.

Enumeration

The enumeration procedure is based on the assumption that a randomly selected sample produces consistent estimates of the aggregate volume or share for choosing a particular alternative. A sample of individuals is drawn, each assumed to represent a group of individuals or households with identical observable characteristics. Probabilities are estimated against each sample data, leading to predictions for average travel demand for individuals or households in each group. These predictions are then aggregated by weighting each sample member according to the corresponding size of the group. In enumeration, the accuracy of the prediction depends on the sample size: the larger the sample size relative to the population of the group, the less will be the sampling error. An example of the enumeration method can be found in the ITS/BTCE transport model (box 3.2).

Enumeration procedures are widely used in urban transport analysis. The major difference between classification and enumeration methods is that the former uses population data and the latter uses sample data. Use of sample data introduces additional errors, namely sampling errors.

Integration

It is also possible to use an integration procedure that weights the disaggregate choice probability estimates by the probability density of the determining explanatory variables. The possible aggregation error caused by this procedure depends on how accurately the distributions are represented. A common practice is to assume a multi-variate normal distribution. The complexity of the process depends on the number of variables over which the integration is taken.

Statistical differentials

The statistical differential approach stipulates that the aggregate shares choosing alternatives are functions of the moments of the joint distribution of exogenous variables. It involves linearising the disaggregate probability function around the mean variable values using a Taylor series expansion in first and higher moments and taking the expectations over the population distributions of the exogenous variables to obtain aggregate model outputs.

The types of aggregation procedures discussed above are useful in developing linkages between disaggregate level models and aggregate level forecasts. The key to success in empirical applications is to identify an aggregation procedure that strikes the right balance between predictions of exogenous variables, computational effort, and magnitude of aggregation errors.

BOX 3.2 ITS/BTCE TRANSPORT MODEL

The Institute of Transport Studies / Bureau of Transport and Communications Economics (ITS/BTCE) model represents urban household travel behaviour in six capital cities in Australia. Developed under contract to the BTCE by a team led by Professor David Hensher from the University of Sydney, the model can be used to evaluate the effects of a wide range of policies on household travel, and travel-related choices.

The ITS/BTCE model consists of a set of inter-related sub-models describing household decisions about choice of residential location, dwelling type, workplace location, number and types of vehicles, commuter mode and time of travel. Choices are arranged into two hierarchical decision trees with residential location choice and fleet size choice at the top. Both sub-models are linked to a sub-model that calculates the continuous variable of vehicle-kilometres travelled (VKT).

A nested logit model is applied to estimate hierarchical choices using survey data collected in 1994 in the six capital cities. The model determines the probability of making a particular choice based on household characteristics and alternative attributes, as well as on other related decisions through inclusive value variables.

Consumer surplus is evaluated at the top level of the residential location choice sub-model, and the fleet size choice sub-model using equation 3.12. Changes in consumer surplus resulting from changes in policies (such as fare reduction for urban public transport, and carbon taxes) are derived from the random utility model as the difference between the value of the 'logsum' before and after the policy change (equation 3.13). The logsum is converted to a dollar measure of consumer surplus by multiplying it by the marginal utility of income, which is equal to the inverse of the parameter estimate associated with the appropriate cost variable in each model.

In estimating total travel demand, the ITS/BTCE model uses an aggregation method similar to the 'enumeration' procedure. 'Synthetic households' are used to represent the actual households in a city. A synthetic household is described in terms of several socioeconomic characteristics. The behaviour of each synthetic household approximates the behaviour of the group of households that have these characteristics. A weight is given to each synthetic household so the size of each group can be estimated. The demand for the entire city can be calculated by summing the demand of all groups of households. Hensher (1993) describes the procedure in more detail.

Source BTCE (1996).

ADVANTAGES OVER CONVENTIONAL MODELS

Disaggregate behavioural travel demand models have a number of advantages over conventional travel demand models. First, they have a sound theoretical foundation, which provides a basis for modelling travellers' behaviour in a consistent manner.

They are also flexible in the sense that feedbacks are allowed among decisions that previously were treated as strictly sequential (Small 1992).

Furthermore, behavioural demand models are policy-sensitive. Direct- and cross-elasticities can be used to simulate the effect of changes in both price-based and non price-based policy variables.

A fourth advantage is that welfare implications of policy changes can be evaluated within the system. This feature has proved to be extremely useful in modern transport analysis.

While behavioural demand models have certain advantages over the conventional approach, they do suffer from a number of limitations. One key limitation is that they tend to concentrate on a particular part of the transport system by focusing on the demand side ‘story’. The supply of the transport system is not explicitly represented in these models, thereby limiting their usefulness in system planning.

Another key limitation of behavioural models is their significant data requirements for forecasting. This shortcoming can be overcome to some extent through aggregation, but the processes involved are not trivial.

CHAPTER 4 LINKED URBAN LAND USE–TRANSPORT MODELS

Historically, theories related to land use and the transport system have developed in relative isolation from each other, although their relationship has been discussed quite extensively in literature for many years (de la Barra 1989).

In earlier work, transport research treated land use variables as being exogenous to the transport system. Since 1970, however, attempts have been made to combine land use and traditional four-step transport models, resulting in the first generation of combined models. Examples of such models include the SELNEC model (Wilson *et al*. 1969; SELNEC, 1971, 1972) and the works of Putman (1973, 1975a, b, c) and Echenique *et al*. (1973). The main characteristic of these models is that the transport model no longer treats land use variables as exogenous and the feedback from the transport system to the land use pattern is explicitly recognised. De la Barra (1989) refers to these types of models as linked land use–transport models.

Figure 4.1 shows the general structure of a typical linked land use–transport model. The land use model in the linked system is mainly concerned with locating 'entities' such as population/housing and employment/workplaces of various kinds to spatial zones. A typical modelling approach involves the use of spatial interaction models based on entropy maximisation (box 4.1). Within such models, the notion of locational accessibility to opportunities plays a central role in the allocation process (box 4.2).

Spatial interaction models incorporating the concept of locational accessibility can be used to simulate relationships between the place of work and the place of residence, and between the place of residence and the place of service activities (such as shops, recreation facilities, education). The outputs of land use modelling are used as inputs to the traditional transport model consisting of trip generation, trip distribution, modal split and route assignment. Travel time and generalised costs are calculated by mapping travel demand on the networks which have capacity constraints.

From the generalised cost calculations, two main feedbacks are incorporated into the model structure: one goes back up to the trip distribution stage in an instantaneous way, and the other back to the location of activities in a lagged

BOX 4.1 SPATIAL ACTIVITY LOCATION MODELS

Spatial activity location models are concerned with representing people's locational decisions in terms of where to live given the place of work, or where to work given the place of residence. These models can be based on the theory of entropy, or on the theory of random utility. Both lead to the same model specification.

Application of entropy maximisation methods to the problem of residential allocation or employment allocation involves use of production- or attraction-constrained spatial interaction models. Take the employment allocation model as an example.

Assume that the number of residents (R_i) is known in each origin zone i. In this case, an origin-constrained spatial interaction model is chosen to allocate residents from zone i to workplaces in zone j. Such a model could take the form:

$$E_{ij} = R_i \mu A_i w_j \exp(-\beta c_{ij}) \tag{4.1}$$

where

$$A_i = \left[\sum_j w_j \exp(-\beta c_{ij}) \right]^{-1} \tag{4.2}$$

and where E_{ij} is the number of residents in i that are allocated jobs in zone j, or a residence-work flow matrix; w_j is a measure of attractiveness of zone j which in this case could be represented by the number of job opportunities; and μ is an employment-to-population ratio. Term A_i ensures that the correct number of jobs is allocated to zone i, ie. $\sum_j E_{ij} = R_i \mu$. Parameter β regulates the effect of transport costs on the distribution of jobs.

The intuition of equation 4.1 is clear. The number of workplaces allocated to zone j from places of residence in zone i is positively related to the attractiveness of zone j and negatively related to the cost of accessing zone j from zone i. It is interesting to note that equation 4.1 is equivalent to a production-constrained model for trip distribution for work purposes.

Alternatively, the employment location model can be derived from a random utility model. Such a model can be based on equation 3.5, and takes the form:

$$E_{ij} = R_i \mu * \frac{\exp(V_{ij})}{\sum_i \exp(V_{ij})} \tag{4.3}$$

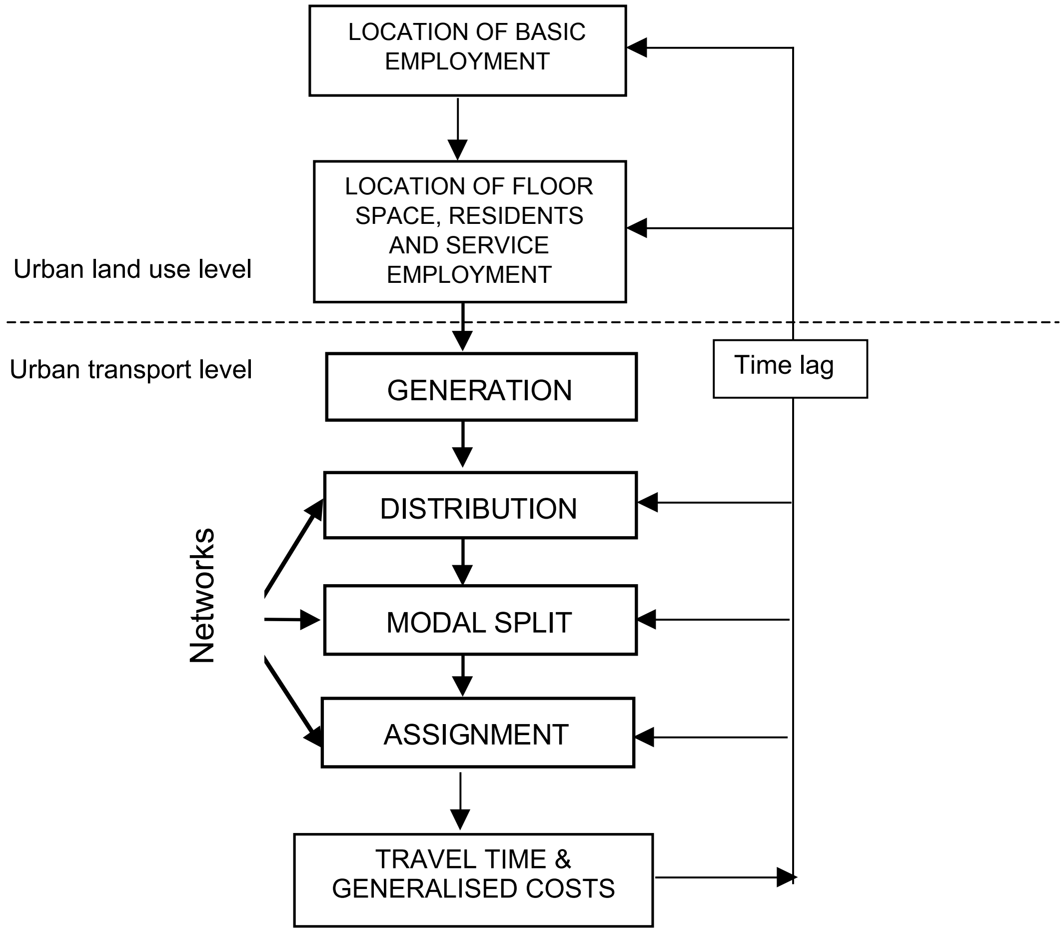
where V_{ij} is a multi-variate utility function. If V_{ij} is set equal to $\ln w_j - \beta c_{ij}$, then

$$E_{ij} = R_i \mu * \frac{w_j \exp(-\beta c_{ij})}{\sum_j w_j \exp(-\beta c_{ij})} \tag{4.4}$$

which is essentially the same as equation 4.1.

Models derived from the random utility theory provide a richer interpretation of the allocation process. The usual elasticity properties and welfare indicators can be derived for policy analysis.

FIGURE 4.1 SCHEMATIC DIAGRAM OF A LINKED LAND USE–TRANSPORT MODEL



Source Based on figure 7.1 of de la Barra (1989), p. 115.

fashion through the notion of accessibility. Inclusion of feedback mechanisms in the linked system is an important improvement over the conventional four-step models which assume a uni-directional causal relationship.

Nevertheless, the linked land use and transport model suffers from a key weakness. The model still treats travel demand at the trip generation stage as being inelastic (unresponsive to) with respect to travel cost. While this may be true for journey-to-work trips, the assumption can not be applied to other types of trips such as shopping and recreational journeys. The demand for these trips is likely to be cost-sensitive.

De la Barra (1989) discusses other problems associated with the linked approach. One is the redundancy of the trip distribution model in the transport system. For instance, in the residential sub-model of the land use system, the relationship between place of work and place of residence is usually simulated by a spatial interaction model, resulting in a residence-work flow matrix. This matrix is similar to, and, if the trip rate (trips per person or per household for a specific period of time) is equal to one, identical to the residence-work trip matrix. It is, therefore, argued that the trip distribution model is redundant, and could be eliminated from the model structure without loss.

Another problem is the use of a simple average method in the calculation of generalised composite (mean) costs. This problem may have been a result of inconsistent modelling techniques used to represent various parts of the system. De la Barra (1989) has shown there is a fallacy in the use of average cost for evaluating user benefits or projects. He argues that the correct formula for calculating generalised composite costs, for example at the origin-destination level, should be based on equation 3.12 and take the form:

$$c_{ij} = \frac{1}{\beta}[\quad {}_{k} \exp(-\beta c_{ij}^{k})] \qquad (4.5)$$

where k refers to all available modes and β is the cost sensitivity parameter which can be estimated from a random-utility-based logit model. Equation 4.5 can be interpreted as the average disutility perceived by travellers, or composite costs. The correct specification of composite costs is important, because it forms a basis for deriving user's surplus indicators which can be used in the evaluation of alternative policies.

BOX 4.2 ACCESSIBILITY

'Accessibility is the concept which combines the geographical arrangement of land use and the transport that serves these land uses' (Black 1981, p. 23).

A high degree of accessibility implies that many land use activities are located close to each other and transport connections are good. Low accessibility results from the wide dispersion of activities and poor transport connections.

Because of unevenly distributed land use activities and transport cost differentials for each pair of spatial zones, accessibility indices are likely to differ from one zone to another.

A commonly applied formula for estimating the accessibility index for a particular zone takes the form:

$$h_i = \quad {}_{j} h_{ij} = \quad w_j^{\alpha} f(c_{ij}) \qquad (4.6)$$

where h_i is the total accessibility of zone i to any nominated activity in all destination zones j; w_j is a measure of attractiveness of a potential destination (or attraction) zone j; α is an economies-of-scale parameter ($0 \leq \alpha \leq 1$); and $f(c_{ij})$ is distance-based cost delay function such as equation 2.9 or 2.10.

Equation 4.6 is closely connected to equation 4.4 in box 4.1. If $\alpha = 1$ and the cost function in the form of equation 2.9 is adopted, the numerator on the right-hand side of equation 4.4 is equal to h_{ij} and denominator to $\quad {}_{j} h_{ij}$. We may therefore view the location-allocation process as a problem of relative accessibility.

CHAPTER 5 INTEGRATED URBAN LAND USE–TRANSPORT MODELS

Integrated urban land use–transport models have gained popularity since the 1980s. Their development represents increasing recognition of intricate connections between land use and transport systems, and of the corresponding need to model these systems in a fully integrated way.

Box 5.1 lists some integrated and empirically applied land use–transport models developed over the past 15 years or so. Most of these models were reported by the International Study Group on Land Use–Transport Interaction (ISGLUTI) (Webster *et al.* 1988). The ISGLUTI study, coordinated through the British Transport and Road Research Laboratory, carried out comparisons of nine different land use–transport models, representing an important milestone in research on integrated models. Other models included in box 5.1 represent more recent developments, some of which were a continuation of the ISGLUTI study.

Although the models listed in box 5.1 are essentially integrated land use–transport models, they involve the use of a variety of theories and modelling techniques. Theories drawn on by these models include microeconomic, entropy/information, random utility, time geography, economic base (box 5.2) and welfare economics theory. Modelling techniques include spatial interaction, random utility, input-output modelling and mathematical programming.

However, at the heart of the more recent models is the combination of the entropy maximisation and locational accessibility premises, that are the basis of spatial interaction theory, with economically rational notions of utility maximisation and consumer choice (Southworth 1995). There has been a tendency to treat spatial interactions as spatial choice problems, which can be addressed by a consistent framework such as the one based on random utility theory.

Integrated land use–transport models can be classified into two fairly distinct groups: predictive models and 'optimising' (or normative) models. Predictive models are based on a set of behavioural relationships. They are concerned with explaining the changing patterns of the land use and transport systems, and with predicting or assessing the impacts of a change in exogenous variables or in policies imposed on these systems. Optimising models aim to map out those land use configurations which will optimise some community objective or a set

BOX 5.1 SOME INTEGRATED AND EMPIRICALLY APPLIED LAND USE – TRANSPORT MODELS

Model	Useful references	Example urban studies
AMERSFOORT*	Floor and de Jong (1981)	Amersfoort, Utrecht, Netherlands; Leeds, UK
BOYCE, *et al.*	Boyce, Tatineni & Zhang (1992) Boyce, Lupa, Tatineni & He (1993)	Chicago, US
CALUTAS*	Nakamura *et al.* (1983)	Tokyo, Nagoya, Okayama, Japan
CATLAS/NYSIM/ METROSIM	Anas (1983b), Anas & Duann (1986), Anas (1992, 1994)	Chicago, New York, US
DORTMUND*	Wegener (1982a, b; 1986, 1995)	Dortmund, Germany
KIM	Kim (1989)	Chicago, US
ITLUP*	Putman (1983, 1991, 1996)	San Francisco, Los Angeles, Houston, Dallas, Portland, others
LILT*	Mackett (1983, 1990a, 1991a, b)	Leeds, England; Dortmund, Germany; Tokyo, Japan
MASTER	Mackett (1990b, c)	Leeds, England
MEPLAN*	Echenique (1985) Hunt & Simmonds (1993) Hunt (1993, 1994)	Bilbao, Spain; Sao Paulo, Brazil; Santiago, Chile; Naples, Italy; others, US
OSAKA*	Amano *et al.* (1985)	Osaka, Japan
POLIS	Prastacos (1986a, b)	San Francisco Bay area
PSCOG	Watterson (1993)	Puget Sound, Washington
TRANSLOC*	Boyce & Lundqvist (1987) Lundqvist (1989)	Stockholm, Sweden
TOPAZ*	Brotchie *et al.* (1980) Dickey and Leiner (1983) Sharpe (1978, 1980, 1982)	Melbourne, Darwin, Australia; Prince William Co. Virginia; others
HAMILTON	Anderson *et al.* (1994); Kanaroglou, *et al.* (1994)	Hamilton, Canada
TRANUS	de la Barra (1989)	Caracas, La Victoria, Venezuela

* indicates participation in the International Study Group on Land Use–Transportation Interaction (ISGLUTI) study.

Source Southworth (1995); updated by BTE.

of objectives. They are designed to evaluate a particular policy or a set of policies in terms of its effect on an objective function (Webster *et al.* 1988). Most of the models listed in box 5.1 are of a predictive nature (except TRANSLOC and TOPAZ), although the distinction is less clear in practical applications because of the incorporation of some behavioural elements in the optimising models.

Depending on the purposes and perhaps computational considerations, the details of representation vary considerably from one model to another, with respect to the land use system or the transport system. Models whose primary focus is on the transport system tend to have a more detailed representation of that system, while leaving the land use system illustrated in a highly aggregated way. The converse applies when the land use system is the focus of attention.

In what follows, a general framework is presented for integrating the land use and transport systems. Because of the great variety of concepts and techniques used in existing operational models, it is not possible to cover all aspects pertinent to the integrated approach. Webster et al. (1988) and Southworth (1995) provide a more detailed and systematic review of integrated land use–transport models. The purpose here is only to provide a general description of a typical integrated land use–transport model. However, the mathematical representation of the relationships included in appendix I, has a strong bias towards the behavioural modelling approach that reflects trends in more recent developments (for example, DORTMUND, MEPLAN and TRANUS). Discussion will focus on predictive models with greater emphasis on transport components.

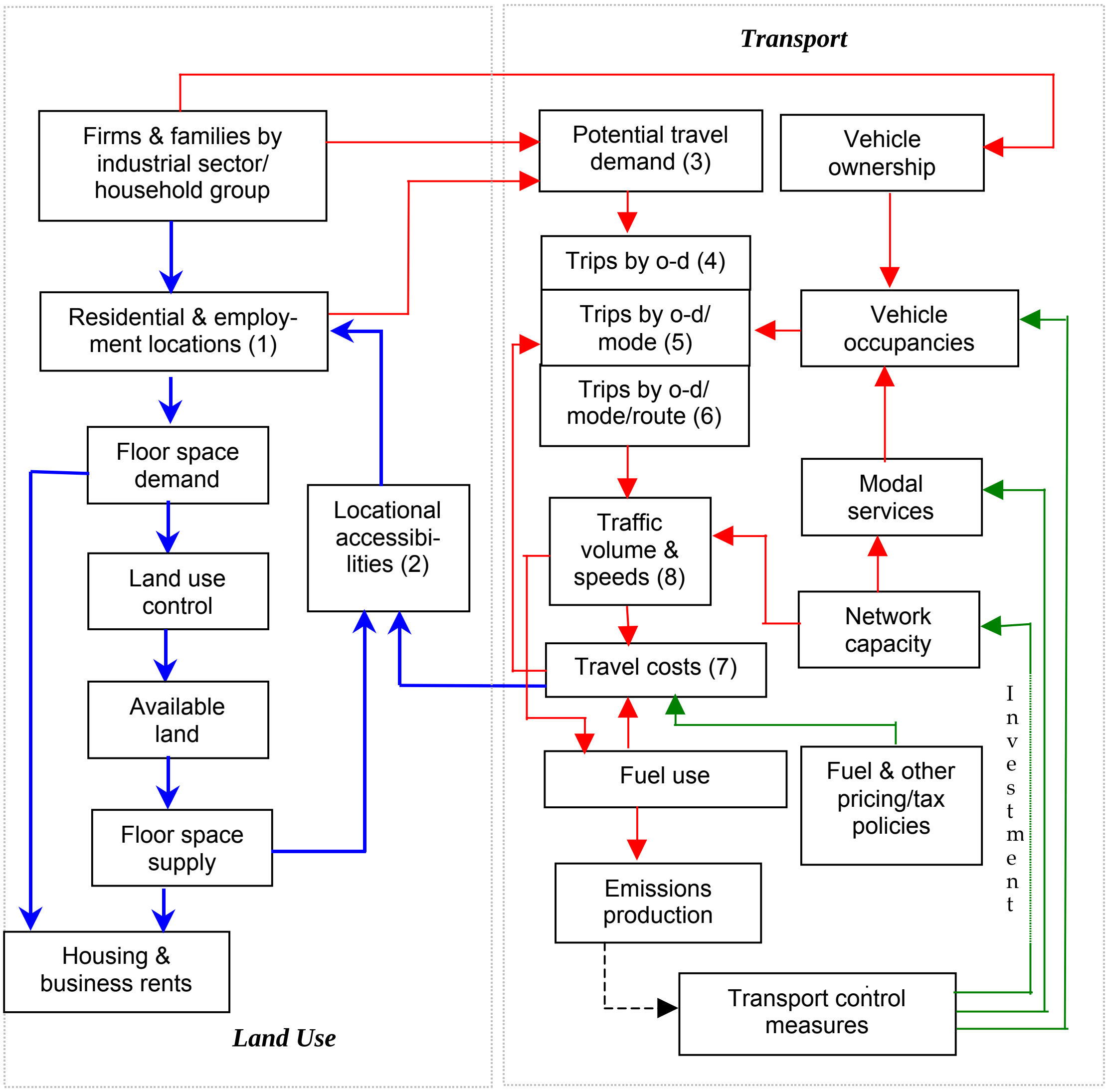
INTEGRATING THE TRANSPORT SYSTEM WITH THE LAND USE SYSTEM

Figure 5.1 provides a schematic representation of an integrated urban land use–transport model. The model consists of two sub-models: one for the land use system and the other for transport. The representation of the land use system is rather simplified, focusing largely on components that interact directly with the transport system. Land use sub-models are intended to explain how spatial choices are made for residential and employment locations. These are basically stipulated as a function of, among other things, locational accessibilities, which in turn depend on zonal attractiveness and travel costs.

The spatial distributions of residents and firms are assumed to create major demand for travel, which drives the development of the transport system. The interplay of demand with supply through transport costs, forms the nucleus of interconnected causes and effects within the transport system.

The land use and transport systems are integrated through a mechanism of feedbacks between the two systems. The land use system supplies the transport system with estimates of the location and volume of travel generators. The transport system affects the land use system through the notion of accessibility, often in a temporally lagged manner. As an integral part of such accessibility, changes in travel costs become part of the mechanism used to relocate labour, residence, and other urban economic activities.

FIGURE 5.1 SCHEMATIC REPRESENTATION OF AN INTEGRATED URBAN LAND USE–TRANSPORT MODEL



Source Adapted from Southworth (1995).

REPRESENTATION OF THE LAND USE SYSTEM

The term 'urban land use' means the spatial distribution or geographical pattern of city function — residential areas, industrial and commercial areas, and the space set aside for government, institutional and leisure functions (Blunden and Black 1984). Three key modules are generally modelled for the land use system: spatial demand for land/floorspace generated from people's choice of residential and employment locations; supply of land/floorspace which is usually controlled by the land use planning authority on the basis of demand; and prices that balance supply and demand. The interaction between these three modules could be explored in detail, but the discussion here focuses on residential and employment locations and locational accessibilities, as these are the two most important interfaces with the transport system.

Residential and employment locations

Residential and employment locations can be conveniently handled within the framework proposed by Lowry (1964), although a great variety of other mechanisms can be used to locate population and employment. The main thrust of the Lowry approach is to decompose total employment into basic (mainly manufacturing) and non-basic (mainly services) employment (box 5.2 and appendix I). Basic employment in each zone — assumed to be exogenously given — together with endogenously determined services employment, are used to estimate the location of residents which in turn is used to predict location of services employment.

BOX 5.2 THE LOWRY MODEL

The Lowry (1964) model is based on two theories: the economic base, and spatial interaction. The economic base concept assumes that an urban economy may be divided into two sectors: a basic sector which produces goods for consumption outside a defined urban area; and a non-basic (or services) sector whose outputs are consumed within the confines of the city. The theory further assumes that the basic sector is the key to the city's growth, and that expansion in the basic sector induces growth in the non-basic sector. In the context of urban modelling, application of the economic base theory implies that the location of basic employment determines the number and location of services employment and population.

In Lowry's original work, residential and employment location allocation was performed within the framework of transport-type gravity models. These models were used to allocate population/households to zones, depending on the place of work. The location of employment in basic industry was specified exogenously, but the location of employment in the services industry was linked directly to the endogenously determined location of population/households.

Lowry's framework has been extended and systematised by Wilson (1970) who suggested four types of entropy models to fit four types of behaviour:

1. Households seeking both a residence and a job: unconstrained model.
2. Households with a residence and seeking a job: production-constrained model.
3. Households with a job and seeking a residence: attraction-constrained model.
4. Households seeking neither residence nor job: doubly constrained model.

The population/household location parts of the Lowry model belong to type 3. Type 4 is not a location model, but it establishes the consistency of the residence-work trip matrix and can generate accessibility indicators for residential zones (Webster *et al.* 1988).

The most used successors to the Lowry model are the Disaggregated Residential Allocation Model (DRAM) and the Employment Allocation Model (EMPAL) in the Integrated Transport Land-Use Package (ITLUP) developed by Putman (Putman 1983, 1991).

In the residential model, population/households are allocated to places of residence in zone i from their places of work in zone j (basic employment plus non-basic employment). Two factors are usually taken into account in the allocation process: zonal attractiveness, and the distance between places of work and places of residence, or the travel cost. Zonal attractiveness can be represented by the availability of floor space for residential use, or housing prices or rents. In some models, zonal attractiveness also includes other characteristics in terms of availability of schools, shops, health services, leisure and recreation facilities. Zonal attractiveness and cost variables usually enter the model in a lagged fashion, data permitting.

Services employment in the Lowry model is linked directly to the location of population/households. This implies that services employment is relatively mobile and is located either contemporaneously with, or soon after, the location of population/households (Webster *et al*. 1988). The major determinants of the location of services employment are accessibility to consumers, and rental costs for services sites.

The Lowry model follows an iterative procedure. It first estimates the location of population/households based on total employment, of which services employment is set to zero for the first iteration. It then proceeds with the location allocation of services employment based on the previously determined location of population/households. The estimated location of services employment will then be added to basic employment to form the basis for the next iteration. In each iteration, a number of residents and services employees are added, however, this number gets smaller and converges to zero after a reasonable number of iterations.

In earlier applications of the Lowry model, interpretation of the above residential and employment locations was based on the entropy maximisation theory. But given that there is a direct connection between spatial interaction models and utility maximising models (box 4.1), interpretation based on the discrete choice theory is preferred.

Accessibility

Accessibility is an important interface between the land use and transport systems. The accessibility of a zone is a measure that combines the convenience with which land-use activities are located in relation to that zone, and the ease or difficulty of reaching these activities via the transport network (Black 1981). Accessibility indicators can be interpreted as attributes of zones, and used as the basis for invoking a random-utility-based location choice model.

In addition to being able to explain the land use pattern, zonal accessibility is an important indicator in its own right. In some studies, accessibility is used as a measure of user benefit to gauge the welfare implications of investment projects, and land use and transport policies.

REPRESENTATION OF THE TRANSPORT SYSTEM

The general structure of transport sub-models in the integrated approach is more or less similar to a standard four-step transport model or linked model. However, there are some important differences, which represent an improvement over the non-integrated approaches. At least three major differences can be identified, although they do not apply to all the integrated models listed in box 5.1. First, trip generation is performed for each zonal pair rather than at zonal aggregate level; second, travel demand is made responsive to changes in travel costs for some types of trips at the trip generation process; and third, a consistent measure based on discrete choice theory is used to estimate user benefits or generalised composite costs.

Potential travel demand

Travel is a derived demand. It is normally undertaken to facilitate a complex and spatially varied set of activities such as work, shopping, recreation and home life (Small 1992). This observation allows demand for transport to be calculated directly from the choices or interactions predicted by the spatial economic system defined within the land use model.

The notion of potential travel demand for work trips at the O-D pair level is equivalent to (or, if the trip rate is one, equal to) the residence-work flow matrix derived from the residential model. The use of information from the land use system to approximate interzonal travel demand avoids possible inconsistencies between the residence-work flow matrix and the distribution matrix for work trips. It also allows fuller exploitation of land use data which are usually more readily available from enumeration censuses.

To derive home-based potential demand for non-work-related trips, origin-constrained spatial choice models can be used, incorporating relevant attractiveness variables for each type of trip. For instance, the pattern of the residence–shop flow matrix depends on the number and quality of shops in the destination zones as well as the cost for accessing these zones. Estimation of potential demand for non-work-related trips could be much more tedious as trip rates cannot be assumed. Separate travel surveys need to be undertaken to estimate trip rates for various purposes.

Trip generation

Trip generation involves use of a function that transforms potential travel demand for each zone pair into actual trips, taking into consideration generalised costs of travel. Such a function could take the form:

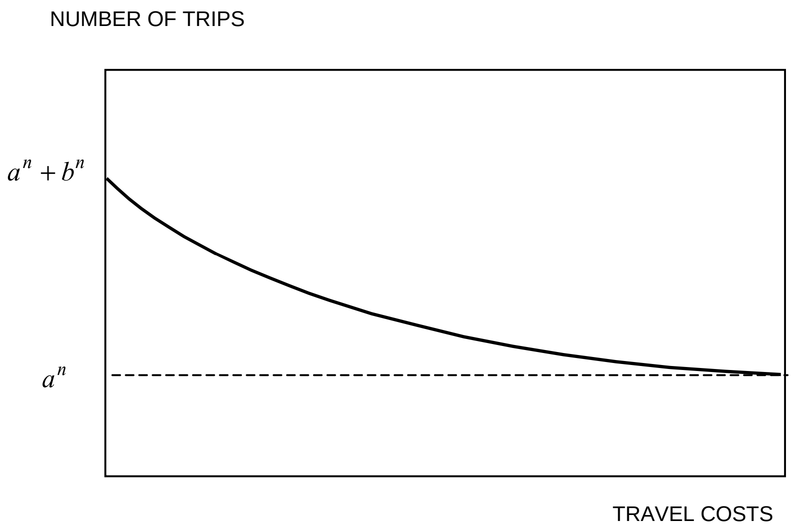
$$T_{ij}^n = Q_{ij}^n[a^n + b^n \exp(-\beta^n c_{ij}^n)] \qquad (5.1)$$

where T_{ij}^n is the total number of trips (or number of trips per household or person) generated from zone i to zone j by activity n; Q_{ij}^n are the functional flows produced by the land use activity model, or potential travel demand.

Equation 5.1 is a downward-sloping curve with a^n being the minimum number of trips that must be performed and $a^n + b^n$ being the maximum. The number of trips decays from the maximum exponentially, with a slope regulated by β^n, as the generalised composite cost c_{ij}^n increases (figure 5.2).

In empirical studies, β^n is usually assumed to be zero (inelastic) for work and school trips but to be greater than than zero for other types of trips such as shopping and recreational trips. If β^n is zero, then the maximum number of trips ($a^n + b^n$) will be made, irrespective of cost.

FIGURE 5.2 TRAVEL DEMAND AND TRAVEL COSTS



Source de la Barra (1989), p. 130.

It should be noted that this treatment of the relationship between travel demand and cost is still quite rudimentary, especially when the model is estimated against cross-sectional data. A truly cost-sensitive travel demand model must also rely on time series or panel data which incorporate information on changes in relative prices of 'transport' and other goods.

Mode choice

Mode choice is mainly concerned with estimating the proportion of trips by each mode for each zone pair. A commonly applied approach is to use a multinomial logit model which expresses mode choice as a function of the disutility or cost of using a particular mode relative to a competing mode.

In some circumstances, because of attribute correlation among various modes, a hierarchical logit model may have to be used to tackle the well-known 'red-bus-blue-bus' conundrum (Mayberry 1970). For example, if bus travel is chosen in preference to car travel, a second choice may follow between, say, red bus and blue bus. Use of the nested approach would effectively help to avoid possible traps in the application of logit models.

Route choice

The route choice model assigns the estimated trips by mode to the individual links of transport networks by determining which route is likely to be followed through the network from one link to the next. Chapter 3 mentioned three types of assignment models. For consistency, a multinomial logit model is again considered here as being appropriate. De la Barra and Perez (1986) proposed an algorithm for such a logit model incorporating capacity constraints. The iterative procedure is described as follows:

1. Compute the first n shortest paths connecting each origin-destination pair, mode, and user type;
2. Assign trips to paths using a multinomial logit model;
3. Once all link loads have been calculated, impose capacity restrictions to calculate the resulting restricted flow speeds; and
4. Finish if convergence has been achieved, otherwise go back to step 1.

While this algorithm meets some important requirements such as having a symmetry property, it is subject to general criticism levied against multinomial logit-based models of potential attribute correlation. In fact, this problem could be even more severe than in the case of mode choice in an operational sense, because the number of routes can be much larger than that of mode options (box 5.3). Therefore, considerable effort is required to overcome the 'red-bus-blue-bus' problem through the hierarchical structuring of models.

Generalised costs and consumers' surplus

The concept of generalised cost involves three elements: out-of-pocket expenses (M), travel time (T) and the value of time (v) (Black 1981). Expressed in monetary units, generalised costs are:

$$C_g = M + vT \tag{5.2}$$

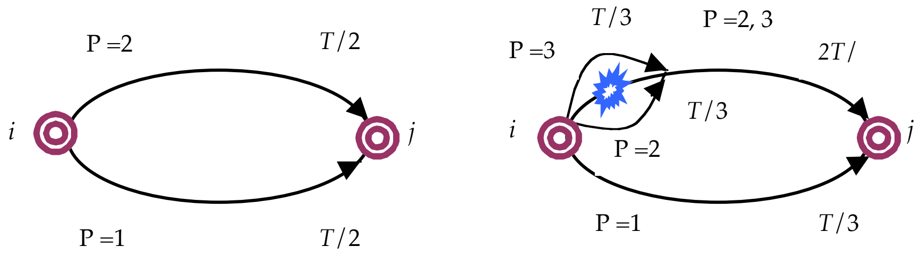
The exchange rate between money and time is a contentious issue. Value of time varies between individuals, and according to the nature of trips taken. A report by BTE (forthcoming) reviews approaches in the valuation of savings in time for various kinds of trips, such as business and non-business travel, and freight transport.

Since C_g are all user-dependent costs, the result is a variable c_{ij}^{nkpl} representing the cost of travel by user type n from i to j by mode k and link l of path p. The total accumulated cost along a path can be obtained by aggregating over l, that is:

BOX 5.3 ATTRIBUTE CORRELATION: 'HOLE IN THE ROAD'

De la Barra (1989) used the 'hole on the road' problem to explain attribute correlation in assignment models. In figure (a), zone i and zone j are linked by two alternative paths (p = 1, 2) which, for simplicity, are assumed to have equal travel costs. In this case, the assignment model based on the logit specification will determine that travellers will be indifferent as to which of the two they should choose, giving equal probabilities for each path. If we denote T as the total volume of traffic between zone i and zone j, then each path will carry $T/2$.

HOLE IN THE ROAD



(a) (b)

One day a big hole opens in path 2 and the construction of two diversions is carried out to avoid it. This creates the new situation depicted in figure (b). Under the new condition, it would not be unreasonable to assume that the each diversion would add a negligible cost, but it would be a mistake to change path options from 2 to 3. The reason for this is that paths 2 and 3 are highly correlated, hence they should be considered as a single option by users. Failure to recognise this would lead to an erroneous result — the logit assignment model would predict T/3 for each path and thus 2T/3 for the problematic road.

The problem of attribute correlation can be solved through the hierarchical structuring of models. In the above example, choices should be first analysed between paths 1 and 2 (figure (a)) and then between the two sub-paths under path 2 (figure 5.3 (b)). Under this structure, probabilities would be revealed as $T/2$ for each path at the first level and $T/4$ (T/2*1/2) for each sub-path at the second level.

$$c_{ij}^{nkp} = {}_{l}\, c_{ij}^{nkpl} \qquad (5.3)$$

Travel time on a particular path is affected by the supply of, and demand for, transport on that path. Specifically, travel time is a function of the volume-capacity ratio. This relationship can be expressed, for example, in the general polynomial form (Black 1981, p. 64):

$$T_Q = T_0[1 + \alpha(\frac{Q}{Q_{\max}})^{\eta}] \tag{5.4}$$

where T_Q is travel time at traffic flow Q; T_0 is 'zero-flow' travel time (free speed); Q is traffic flow, vehicles per hour; Q_{max} is 'practical capacity' which is defined as three quarters of saturation level; and α, η are parameters.

There are also other specifications that describe the traffic flow-dependent travel time relationship. Choice of a particular specification depends largely on empirical data.

Calculations of generalised composite costs must be performed according to the formula specified in equation 3.12 and proceeding along the decision chain from the link level upwards. This involves aggregation of composite costs over path p for input into the mode choice model and further over k for input into the trip generation model.

Estimation of user benefit from a nested logit model is straightforward, athough care needs to be taken to avoid double-counting. The formula for calculating changes in user benefit should be along the lines of equation 3.13 and be performed only at the mode choice level. Such a formula takes the form:

$$\Delta W^n = -\frac{1}{\beta^n}\ln[\frac{\sum_k T_{ij}^{nk}\exp(-\beta^n c_{ij}^{nk})(2)}{\sum_k T_{ij}^{nk}\exp(-\beta^n c_{ij}^{nk})(1)}] \tag{5.5}$$

where (2) denotes the scenario being evaluated and (1) the base case against which (2) is being compared. Note that T_{ij}^{nk}, the number of trips by user type n and by mode k, is included in equation 5.5 to account for the difference in the number of trips in each scenario due to elastic trip generation. The set of indicators based on equation 5.5 represent changes in utility per trip made, and must be aggregated to a system-wide single value (de la Barra 1989). This value, when multiplied by the marginal utility of income, gives rise to a dollar measure of consumer surplus (BTCE 1996, p. 389).

APPLICATIONS IN PLANNING AND POLICY ANALYSIS

Integrated land use and transport models such as the one illustrated in figure 5.1 can be used to predict future land use patterns and transport demand by introducing changes in exogenously determined variables such as population and basic employment. Accuracy of prediction, especially for the long run, depends largely on whether urban dynamics is built into the model structure. For most models where calibration is based on cross-sectional data, predictions are likely to be valid only for the short term. Longer-term projection requires that dynamic interaction be fully specified in the model and be supported by longitudinal or panel data (a combination of time-series and cross-sectional data).

Nevertheless, cross-sectionally calibrated models can be still useful in simulating the response of the land use and transport systems to changes in policies in the same way that comparative-static analysis does.

Common policy applications can be classified into three broad categories: regulatory, pricing and investment policies (Webster *et al*. 1988, pp. 136-138). *Regulatory policies* include those that regulate the use of space or time. Examples of space-related transport policies are the reservation of road space for say, 'bus only' or 'high occupancy vehicles only', or pedestrian use and controls on street parking. Examples of time-related policies include time scheduled for public transport, or when a specific traffic arrangement takes effect. *Pricing policies* are those which directly affect the price of land, buildings or transport. Transport policies of this type include, for example, the imposition of fuel or emissions taxes, road tolls, congestion charges, parking charges, and subsidisation of public transport. *Investment policies* are those which directly affect the capacity of the transport network, and hence travel speeds. Determining the level and type of investment in transport infrastructure has been a central concern to transport planners and policy makers. This concern has diminished in developed countries, because of slower growth in traffic demand due to low growth in population, slower increase in incomes, and already high levels of vehicle ownership.

The general framework described in figure 5.1 can be modified or extended to evaluate policies for particular objectives, such as those to limit peripheral urban development, to encourage use of public transport and reduce car dependence, and to conserve resources such as time or energy. The model could also be coupled with a detailed economic module to assess the welfare implications of various policies.

LIMITATIONS

Although used for almost 20 years, integrated urban land use–transport models are still underdeveloped in a number of areas, and are open to further improvement in view of ever challenging tasks faced by transport planners and policy makers. The discussion below draws heavily on Southworth (1995).

In existing studies, the issue of trip chaining has not been addressed effectively. Trip chaining refers to the fact that many trip destinations in urban areas occur within multi-purpose, multi-stop daily travel chains. For example, from home to school to work to school to shop to home. Trip chaining, which is related closely to people's daily activity schedules, has time- and cost-saving effects. Ignoring these effects in current operational models may prevent the design of land use policies which might take advantage of this time- and cost-saving option. A further consequence is that, if a possible destination is removed from the available choice set, this would affect the absolute probabilities calculated within a logit model.

Off-peak travel activity modelling has also been relatively neglected. Peak-hour modelling may be sufficient for transport planners whose primary interest is to reduce actual or potential traffic congestion. But for a city which is congested all day, or for a policy-maker who has other policy interests, such as maximisation of net social benefits or reduction in greenhouse effects caused by road traffic, non-peak travel activity modelling incorporating multiple purposes becomes necessary. Moreover, there is a growing recognition of the need to treat departure time as a choice problem for travellers, rather a simplistic split made by the modeller between peak- and off-peak travel.

There is still a weakness in the treatment of the relationship between trip frequencies and travel costs. As most transport studies are of necessity based on cross-sectional data, the concept of relative price (transport over other goods) cannot be applied. Put another way, a total separability assumption has to be made — meaning that decisions about spending on transport are made separately from decisions about spending on other commodities. This creates a problem for forecasting, because in the long run, price-induced substitution is likely to occur between transport and other household goods.

Urban dynamics has been tackled in a relatively primitive way. Most models employ quasi-dynamic approaches to multi-year forecasting or scenario generation. That is, they move from one period to the next, updating the value of variables as they pass through successive time periods, in accordance with dynamic controlling mechanisms built into the system. There is a need to move gradually towards more behaviourally consistent and true dynamic modelling approaches based on difference or differential equation forms and supported by longitudinal data. To this end, systematic design for data collection is crucial.

Another area of deficiency in current practice is the underdeveloped treatment of urban freight modelling. An explicit, behaviourally realistic decision-making framework is needed in freight movement research.

CHAPTER 6 COMPARATIVE EVALUATION OF TRANSPORT MODELS

Chapters 2-5 reviewed each of four broad types of urban transport models. The evolution of these models from the traditional four-step approach to linked, then to the integrated approach, represents an increasing understanding of the urban transport system and its interaction with the land use system. The development of behavioural demand models based on random utility theory, and the introduction of these models into aggregate analysis adds further to the sophistication and behavioural richness of zone-based urban transport models.

COMPARISON OF FEATURES OF FOUR TYPES OF TRANSPORT MODELS

The types of models of primary interest to policy-makers are those operational at the strategic zonal level, capable of predicting long-term travel demand, capable of explaining changes in land use, involving dynamic interaction between land use and transport, and having the ability to explain competition between different transport modes. The four types of models that have been reviewed differ significantly in their capacity to deliver these desirable characteristics (table 6.1).

TABLE 6.1 COMPARISON OF FOUR BROAD CATEGORIES OF TRANSPORT MODELS

Desirable characteristics	Conventional	Behavioural	Linked	Integrated
Zone-based analysis	Y	QN	Y	Y
Long term-travel demand prediction	N	N	N	QY
Long-term land use prediction	N	N	N	QY
Dynamic interaction	N	N	QY	QY
Mode choice	Y	Y	Y	Y

Notes Y=yes; QY=qualified yes; N=no; QN=qualified no.

Source BTE 1998.

Traditional four-step transport models are based on zonal data. These models are not intended to provide long-term land use predictions, because these are established outside the modelling system. While traditional models may be used for scenario generation — typically over a five-year planning horizon — their ability to make long-term travel predictions is hampered by the lack of dynamic interaction with the land use system. Being cross-sectionally calibrated, most traditional models are static by nature. Mode choice analysis is usually given greater behavioural treatment (ie. use of logit models), but it is modelled as an independent process within the system.

Behavioural travel demand models were initially developed to analyse individual travel behaviour, notably mode choice. However, they rapidly penetrated into zone-based analyses through aggregation, and into analyses of other travel and locational choices. Behavioural demand models in their disaggregate form are not capable of making long-term predictions of travel demands or land use patterns. These models also tend to be static, using cross-sectional data only.

Linked land use–transport models represent early attempts to integrate the land use system with the transport system. These models tend to incorporate, if data are available, a dynamic interaction mechanism whereby transport costs have lagged impacts on the land use pattern. However, the assumed price-inelasticity of travel demand could be a potential source of distortion in travel demand prediction and possibly land use prediction as well.

Integrated land use–transport models possess most of the desirable characteristics. The reason why a 'qualified yes' is given in table 6.1 to the integrated approach for its capacity to deliver long-term prediction and urban dynamics is that this capacity is largely data-dependent. For a single cross-sectional representation of an urban system, the best planning horizon for a forecast may be less than 5 years. Forecasting further into the future requires more sophisticated modelling techniques, supported by panel data.

SUITABILITY FOR POLICY ANALYSIS

Table 6.2 assesses the suitability of various types of transport models in terms of broad classes of policy analysis. For any model to be suitable for policy analysis, it should have behavioural richness, be sensitive to policy changes, and be able to provide economic evaluation of policies.

The traditional four-step approach fails to a large extent on all these counts. Most of the sub-models in the four-step analysis are not derived from a consistent theoretical perspective. The traditional approach is also insensitive to pricing policies, as total travel demand is assumed to be inelastic with respect to travel costs. Moreover, it is unable to provide analysis of welfare implications of policy changes based on sound economic theory. The usefulness of the traditional approach in policy analysis is therefore limited.

TABLE 6.2 SUITABILITY FOR POLICY ANALYSIS

	Conventional	Behavioural	Linked	Integrated
Transport regulations	+	++	+	++
Pricing policies	+	++	+	++
Investment policies	+	-	+	++
Welfare implications	-	++	-	++

Notes - not suitable; + partly suitable; ++ suitable.

Source BTE 1998.

Behavioural travel demand models have an acceptable theoretical base, a capability of responding to a much wider range of policy options and a measure to gauge the welfare implications of policy changes. These models tend to be used to deal with specific aspects of transport systems in greater detail to support policy evaluation, and sometimes for aggregate analysis. Behavioural demand models, if not aggregated at the strategic zonal level, are not very useful in travel demand prediction, which is vital for evaluation of investment policies.

Linked land use–transport models are essentially extensions of the traditional four-step models, and therefore suffer from the same weaknesses.

Integrated land use–transport models represent the state-of-the-art tool for policy analysis. This is particularly true of the more recent models based on the random utility approach throughout the model structure. The particular attraction of these models is that they can not only simulate the effects of changes in various policies, but also provide an economic evaluation of policy changes. The latter is very important when considering the desirability of a particular policy.

It should be noted that the distinction between the four types of models becomes blurred when zone-based analysis is given a behavioural treatment. Integrated models can therefore be regarded as a general framework, the simplification of which could result in, for example, a four-step transport partial model embodying behavioural relationships. This gives flexibility to transport analysts for tailoring the large integrated model to suit particular policy applications.

POSSIBLITIES FOR THE FUTURE

Urban modelling is a complex and costly task. Like any other modelling exercise, it is important to identify planning needs and policy interests before deciding on any particular modelling style.

For a long time, forecasting growth in travel demand and accommodating that growth has been a central concern to transport planners. This is expected to continue to be the case, although the relative importance of this need in planning agendas may decline to some extent. Faced with persistent congestion, pollution, accidents and financial strains in many capital cities in Australia, transport planners and policy makers have now turned their attention to transport demand management. Clearly, there are interests, both at the Commonwealth and state levels, in travel reduction strategies that could lead to less congestion, less pollution and fewer accidents, as well as other desirable outcomes.

The nature of the planning needs and policy interests suggest that the more general framework provided by integrated land use–transport models should be adopted. Determining travel demand, especially for the long run, requires that land use patterns be modelled because of the close interaction with the transport system.

Detailed modelling of land use patterns would also keep options open for exploiting land use policies in the design of travel reduction strategies. Interest in reducing congestion necessitates a very detailed network representation in the transport sub-model, as a congestion-abating effort on a particular link would have strong network effects. Exploring options for cutting greenhouse gas emissions from the road sector requires all types of trip to be analysed — work and non-work trips as well as peak and off-peak hour trips. Concerns over efficiency and equity issues give rise to the need for the model to have a sound measure of welfare derived from a consistent framework such as that provided by random utility theory.

While the underlying framework should be made as general as possible, development of a comprehensive model can be approached incrementally, depending on the priorities of planning needs and policy interests as well as the availability of data. For instance, the Commonwealth Government has an interest in assessing long-term funding requirements for those parts of the national road network which connect strategic locations such as seaports, airports, rail terminals, distribution centres and military sites in metropolitan centres to the National Highway. The network representation of the transport system can be simplified to a considerable degree to satisfy these needs, although some modelling of the land use systems with dynamic interaction is unavoidable. Similarly, if capacity constraints are most manifest in commuting time, an analysis of commuting trips may be sufficient. However, it is important to ensure that the model structure can easily be modified or extended to encompass other types of analysis in the future.

There are tough challenges. They involve the collection of significant amounts of data, accumulating human technical skills in transport modelling, and tackling theoretical and practical issues that have been unanswered in the existing studies. However, if the increasingly complex needs of policy formulation are to be met in a sufficiently rigorous way, the development of improved models is then necessary.

APPENDIX I MATHEMATICAL REPRESENTATION OF AN INTEGRATED URBAN LAND USE–TRANSPORT MODEL

This appendix provides a mathematical account of the integrated urban land use–transport model illustrated in figure 5.1. The derivation of the relationship is based largely on the standard Lowry framework and behavioural choice theory, drawing heavily on the demonstration by de la Barra (1989). The model can be treated as a prototype that could be extended to encompass more complex relationships.

(1) Residential and employment locations

Residential and employment locations can be modelled within the framework of Lowry's (1964) work. The Lowry model defines the urban system as composed of a basic employment sector, a service employment sector and a residential sector. Basic employment in each zone is exogenously determined and is used to estimate the location of residents and service jobs. Other exogenous variables are an accessibility matrix, defined in section 2 of this appendix.

The Lowry model is of sequential nature and intended to be iterative. Expressed in terms of a series of singly constrained spatial interaction models, the calculations are performed as follows:

a) Add basic employment (E_j^b) and service employment (E_j^s) calculated in the last iteration to estimate total employment in zone *j* (E_j) :

$$E_j = E_j^b + E_j^s \qquad \text{(A1)}$$

In the first iteration, the service employment (E_j^s) is set to zero.

b) Allocate residents to zone *i* from work places in *j*:

$$R_{ij} = E_j \mu B_j w_i^{\alpha} \exp(-\beta^r c_{ij}) \qquad \text{(A2)}$$

where

$$B_j = \left[\sum_i w_i^{\alpha} \exp(-\beta^r c_{ij}) \right]^{-1} \qquad \text{(A3)}$$

and where R_{ij} is the number of residents of i that work in j; μ is a population-to-employment ratio; w_i measures the attractiveness of zone i which, in this case, could be represented by the availability of floor space for residential use. Term B_j ensures that the correct number of residents is allocated to zone i, ie. $\sum_i R_{ij} = E_j\mu$. Parameter β^r regulates the effect of transport costs on the distribution of residents: a high value of β^r will result in the population being allocated close to their place of work; if $\beta^r \rightarrow \infty$, all residents will live and work in the same zone; if $\beta^r \rightarrow 0$, residents will locate in proportion to the available land area in equal density.

c) Allocate services employment to zones j from places of residence in zone i:

$$E_{ij}^s = R_i\, s A_i w_j^\alpha \exp(-\beta^s c_{ij}) \tag{A4}$$

where $$A_i = \left[\sum_j w_j^\alpha \exp(-\beta^s c_{ij})\right]^{-1} \tag{A5}$$

and where s represents a service-to-population ratio; w_j is the availability of floor space for commercial use. Term A_i ensures that the correct number of service employees is allocated to zone j, ie. $\sum_j E_{ij}^s = R_i s$. The total number of service employees in zone j can be obtained:

$$E_j^s = \sum_i E_{ij}^s \tag{A6}$$

Calculation then returns to step (a), where service employment is added back to exogenous basic employment. At each iteration, a number of residents and service employment are added, however, this number becomes progressively smaller, converging after a number of iterations.

The services employment sub-model can be disaggregated by considering different types of services, such as retail shops, education, health and so on. This would form a basis for analysis of travel demand by different purposes.

It should be noted that the original Lowry model was developed in the context of gravity or entropy theory. While more recent models have kept the Lowry-type structure, each sub-model for location allocation tends to be based on behavioural choice theory. This ensures a consistency between behaviour in locational choices and that in travel choices.

(2) Accessibility

The accessibility of a particular zone is an indication of how land-using activities (such as population and employment), are located in relation to that zone and how difficult it is to reach them via the transport network (Black 1981). Total accessibility of zone i to any nominated activity in all destination zones j (including zone i), h_i, is thus a function of land use intensity in zone j (L_j) and the cost of interaction (c_{ij}):

$$h_i = f(L_j, c_{ij}) \tag{A7}$$

The indicator of land use intensity in zone j (L_j) may be interpreted as a measure of attractiveness of zone j which is denoted as w_j. Depending on specific situations, w_j may refer to different indicators. For instance, in the residential model where residents decide where to live, w_j could represent the availability of floor space for residential use; and in the employment location model, where workers decide where to work, it could refer to job opportunities in zone j.

A commonly applied empirical model for estimating the accessibility indicator takes the form:

$$h_i = \sum_j h_{ij} = \sum_j w_j^{\alpha} \exp(-\beta c_{ij}) \tag{A8}$$

where α is a parameter regulating w_j and is sometime called the economies-of-scale parameter. Intuitively, this equation says that the accessibility of a zone depends on how that zone is located in relation to zones of attraction and the cost of interaction between zone i and zone j.

(3) Potential travel demand

Potential travel demands can be derived from locational models at the land use level. In the case of a journey-to-work trip, the potential demand for travel is given by the residence-work flow matrix, ie.:

$$Q_{ij}^{w} = R_{ij} \tag{A9}$$

where Q_{ij}^{w} is the potential demand for work trips at the origin-destination pair level; R_{ij} is defined in (A2).

For trips of other purposes, potential demand can be estimated by using an origin-constrained spatial choice model which has the form:

$$Q_{ij}^{n} = R_i A_i (w_j^n)^{\alpha} \exp(-\beta^n c_{ij}^n) \tag{A10}$$

where

$$A_i = \left[\sum_j (w_j^n)^{\alpha} \exp(-\beta^n c_{ij}^n) \right]^{-1} \tag{A11}$$

and where Q_{ij}^{n} is the potential demand for non-work related trips such as school trips, shopping trips, recreational trips and so on. Corresponding to the n purposes of trips, there is set of variables describing the characteristics of attracting zone j, w_j^n, representing the number of schools, or shops or recreational facilities in zone j. A_i is the balancing factor ensuring that the simulated R_i is equal to endogenously determined R_i. Note that the composite costs, c_{ij}^{n}, need to be estimated for each trip purpose n.

The total potential demand for travel can be obtained by aggregating Q_{ij}^{n} over n plus Q_{ij}^{w}:

$$Q_{ij} = \sum_n Q_{ij}^{n} + Q_{ij}^{w} \tag{A12}$$

where for example, n=1 for school trips, n=2 for shopping trips, n=3 for recreational trips, and n=N for other trips.

(4) Trip generation

Trip generation involves use of a function that transforms potential demand into actual trips, taking into consideration generalised costs of travel. Such a function could take the form:

$$T_{ij}^{n} = Q_{ij}^{n}[a^{n} + b^{n} \exp(-\beta^{n} c_{ij}^{n})] \quad \text{(A13)}$$

where T_{ij}^{n} is the total number of trips (or number of trips per household or person) between zones i and j for n purposes (for simplicity, n includes work trips as well).

The demand function A14 for trips has the usual downward slope of demand curves, with a^n being the minimum number of trips which must be performed and $a^n + b^n$ being the maximum. The number of trips decays exponentially from the maximum, with a slope regulated by β^{n}, as the generalised composite cost c_{ij}^{n} increases.

In empirical studies, β^{n} is usually assumed to be zero (inelastic) for work and school trips but to be greater than zero for other types of trips such as shopping and recreation.

(5) Mode choice

In modal split modelling, the most common practice is to use the multinomial logit model which in this case takes the form:

$$T_{ij}^{nk} = T_{ij}^{n} \frac{\exp(-\beta^{n} c_{ij}^{nk})}{\sum_{k} \exp(-\beta^{n} c_{ij}^{nk})} \quad \text{(A14)}$$

This equation states that the number of trips made in mode k for activity n is the product of the total number of trips and k's share in these trips.

(6) Route choice

A similar multinomial logit model can be applied to route choice modelling, taking the form:

$$T_{ij}^{nkp} = T_{ij}^{nk} \frac{\exp(-\beta^{n} c_{ij}^{nkp})}{\sum_{p} \exp(-\beta^{n} c_{ij}^{nkp})} \quad \text{(A15)}$$

Intuitively, this equation states the number of trips made in mode k via path p for activity n, is a product of the total number of trips made in mode k and the p's share in these trips.

(7) Generalised costs and consumers' surplus

Generalised costs involve three elements: out-of-pocket expenses (M), travel time (T) and the value of time (v) (Black 1981). Expressed in monetary units, generalised costs are:

$$C_g = M + v\,T \qquad \text{(A16)}$$

Since C_g are all user-dependent costs, the result is a variable c_{ij}^{nkpl} representing the cost of travel for activity n from i to j by mode k and link l of path p. The total accumulated cost along a path can be obtained by aggregating over l. That is:

$$c_{ij}^{nkp} = \;{}_{l}\, c_{ij}^{nkpl} \qquad \text{(A17)}$$

Travel time on a particular path is influenced by transport supply and demand on that path. More specifically, it is a function of volume-capacity ratio. Such a relationship can be expressed in the general polynomial form (Black 1981):

$$T_Q = T_0[1 + \alpha(\frac{Q}{Q_{\max}})^{\eta}] \qquad \text{(A18)}$$

where T_Q is travel time at traffic flow Q; T_0 is 'zero-flow' travel time; Q is traffic flow, vehicles per hour; Q_{max} is 'practical capacity' which is defined as three quarters of saturation level; and α, η are parameters.

Calculations of generalised composite costs must be performed according to the formula specified in equation 3.12 and proceeding backwards along the decision chain. The composite costs of travel from origin i to destination j by mode k can be obtained by aggregating c_{ij}^{nkp} over all paths p. This is given as:

$$c_{ij}^{nk} = \frac{1}{\beta^n}\ln[\;{}_{p}\exp(-\beta^n c_{ij}^{nkp})] \qquad \text{(A19)}$$

Similarly, the composite costs of travel from an origin i to destination j can be obtained by aggregating c_{ij}^{nk} over all modes k:

$$c_{ij}^{n} = \frac{1}{\beta^n}\ln[\;{}_{k}\exp(-\beta^n c_{ij}^{nk})] \qquad \text{(A20)}$$

Now the concept of consumers' surplus can be introduced. Changes in consumers' surplus can be calculated at the route choice or mode choice level, but to avoid double-counting, only consumers' surplus at mode choice level needs to be evaluated. The formula for calculating this indicator should follow equation 3.13. In this case it takes the form:

$$\Delta W = -\frac{1}{\beta^n}\ln[\frac{{}_{k}T_{ij}^{nk}\exp(-\beta^n c_{ij}^{nk})(2)}{{}_{k}T_{ij}^{nk}\exp(-\beta^n c_{ij}^{nk})(1)}] \qquad \text{(A21)}$$

where (2) denotes the scenario being evaluated and (1) the base case against which (2) is being compared.

REFERENCES

Abbreviations

ABS	Australian Bureau of Statistics
AGPS	Australian Government Publishing Service
BTCE	Bureau of Transport and Communications Economics
BTE	Bureau of Transport Economics

ABS 1995, *Survey of Motor Vehicle Use*, Catalogue No. 9202.0, AGPS, Canberra.

Amano, K. and Abe, H. 1985, 'An activity location model for the metropolitan area', *Infrastructure Planning Review*, No. 2, pp. 165-172.

Anas, A. 1983a, 'Discrete choice theory, information theory and the multinomial logit and gravity models', *Transport Review*, 17B, pp. 13-23.

Anas, A. 1983b, *The Chicago Area Transportation–Land Use Analysis System*, final report to US Department of Transportation, Washington, D.C.

Anas, A. 1992, *NYSIM (The New York Simulation Model): A Model of Cost–Benefit Analysis of Transportation Projects*, New York Regional Planning Association, New York.

Anas, A. 1994, *METROSIM: a Unified Economic Model of Transportation and Land Use*, pamphlet distributed at the Transportation Model Improvement Program's Land Use Modelling Conference, Dallas, Texas, February 19-21, 1995.

Anas, A. and Duann, L. S. 1986, 'Dynamic forecasting of travel demand, residence location, and land development: policy simulations with the Chicago area transportation–land use analysis system', in *Advances in Urban Systems Modelling*, ed. B. Hutchinson and M. Batty, North-Holland, Amsterdam, pp. 299-322.

Anderson, W. P., Kanaroglou, P. S. and Miller, E. J. 1994, *Integrated Land Use and Transportation Model of Energy and Environmental Analysis*, a report on design and implementation, Department of Geography, McMaster University, Ontario, Canada.

Beimborn, E. A. 1995, *A Transportation Modelling Primer*, Centre for Urban Transportation, University of Wisconsin-Milwaukee, http://www.uwm.edu/Dept/CUTS/primer.htm.

Ben-Akiva, M. E. 1973, *Structure of Passenger Travel Demand Models*, PhD Dissertation, Department of Civil Engineering, MIT, Cambridge, Mass.

Black, J. 1981, *Urban Transport Planning — Theory and Practice*, Croom Helm, London.

Blunden, W. R. and Black, J. A. 1984, *The Land-use Transport System*, 2nd edition, Pergamon Press, Sydney.

Boyce, D. E. and Lundqvist, L. 1987, 'Network equilibrium models of urban location and travel choices: alternative formulations for the Stockholm region', *Papers of the Regional Science Association*, No. 61, pp. 93-104.

Boyce, D. E., Lupa, M. R., Tatineni, M. and He, Y. 1993, *Urban Activity Location and Travel Characteristics: Exploratory Scenario Analyses*, December 6-10, SIG1 Seminar Paper on Environmental Challenges in Land Use Transport Coordination, Blackheath, Australia.

Boyce, D. E., Tatineni, M. and Zhang, Y. 1992, *Scenario Analyses for the Chicago Region with a Sketch Planning Model of Origin–Destination Mode and Route Choice*, final report to Illinois Department of Transportation.

Brotchie, J. F., Dickey, J. W. and Sharpe, R. 1980, 'TOPAZ: planning techniques and applications', *Lecture Notes in Economics and Mathematical Systems Series*, Vol. 180, Springer-Verlag, Berlin, Germany.

BTCE 1996, *Transport and Greenhouse: Costs and Options for Reducing Emissions*, Report 94, AGPS, Canberra.

BTCE 1997, *Roads 2020*, Working Paper 35, Bureau of Transport and Communications Economics, Canberra.

BTE 1998, *Forecasting Light Vehicle Traffic*, Working Paper 38, Bureau of Transport Economics, Canberra.

BTE (forthcoming), *Facts and Furphies in Cost-Benefit Analysis: Transport*, Bureau of Transport Economics, Canberra.

Button, K. J. 1977, *The Economics of Urban Transport*, Saxon House, Farnborough, UK.

Daganzo, C. F. 1980, 'Optimal sampling strategies for statistical models with discrete dependent variables', *Transport Science*, Vol. 14, pp. 324-25.

Daly, A. J. and Ortuzar, J. de D. 1990, 'Forecasting and data aggregation: theory and practice', *Traffic Engineering + Control*, December, pp. 632-643.

Daly, A. J. and Zachary, S. 1978, 'Improved multiple choice models', in *Determinants of Travel Choice*, ed. D. Hensher and M. Q. Dalvi, Saxon House, Westmead.

de la Barra, T. 1989, *Integrated Land Use and Transport Modelling: Decision Chains and Hierarchies*, Cambridge University Press, Cambridge.

de la Barra, T. and Perez, B. 1986, '*Asymmetry in Some Common Assignment Algorithms: the Dispersion Factor Solution*', Environment and Planning, No.13(B), pp. 293-304.

Dickey, J. W, and Leiner, C. 1983, 'Use of TOPAZ for transportation–land use planning in a suburban county', *Transportation Research Record*, No. 931, pp. 20-26.

Echenique, M. H. 1985, 'The use of integrated land use–transportation planning models: the cases of Sao Paulo, Brazil and Bilbao, Spain', in *The Practice of Transportation Planning*, ed. M. Florian, Elsevier, Netherlands.

Echenique, M., Feo, A., Herrera, R. and Riquezes, J. 1973, *A Disaggregated Model of Urban Spatial Structure: Caracas, Working Paper*, No. 8, Cambridge: Land use and Built Form Studies.

Floor, H. and de Jong, T. 1981, 'Testing a disaggregated residential location model with external zones in the Amersfoort region', *Environment and Planning*, No. 13A, pp. 1473-83.

Fotheringham, A. S. and O'Kelly, M. E. 1989, *Spatial Interaction Models: Formulations and Applications*, Kluwer Academic Publishers, London.

Fraser, D. A. S. 1958, *Statistics: An Introduction*, New York: John Wiley.

Friesz, T. L. and Harker, P. T. 1985, 'Freight network equilibrium: a review of the state of the art', in *Analytical Studies in Transport Economics*, ed. A. F. Daughety, Cambridge University Press, London.

Greene, W. H. 1991, *Econometric Analysis*, MacMillan Publishing Company, New York.

Hensher, D. A. 1986, '*Simultaneous Estimation of Hierarchical Logit Mode Choice Models*', Working Paper No. 24, Macquarie University, School of Economic and Financial Studies.

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

Hensher, D. A. 1997, 'Stated preference analysis of travel choices: the state of art', in *Transport Economics: Selected Readings*, eds. T. H. Oum *et al*., Harwood Academic Publishers in cooperation with the Korea Research Foundation for the 21st Century, Australia.

Hunt, J. D. 1993, A Description of the MEPLAN Framework for Land Use and Transport Interaction Modelling, paper presented at the 73rd Annual Transportation Research Board Meetings, January 9-13, Washington, D.C.

Hunt J. D. 1994, *Calibrating the Naples Land Use and Transport Model* (Draft), Department of Civil Engineering, University of Calgary, Calgary, Alberta, Canada.

Hunt, J. D. and Simmonds, D. C. 1993, 'Theory and application of an integrated land use and transport modelling framework', *Environment and Planning*, No. 20B, pp. 221-244.

Jones, P., Dix, M., Clark, M. and Heggie, I. 1983, *Understanding Travel Behaviour*, Gower, Aldershot, UK.

Kanaroglou, P. S., Anderson, W. P., Miller, E. J. and Buliung, R. N. 1994, Estimating Emissions and Fuel Consumption for Automobile Traffic in Hamilton Ontario (Draft), Department of Geography, McMaster University, Ontario, Canada.

Kim, T. J. 1989, *Integrated Urban System Modelling: Theory and Practice,* Martinus Nijhoff, Norwell, Massachusetts.

Koppelman, F. S. 1984, 'The application of disaggregate choice models to travel demand forecasting: issues and methods', in *Transport Planning Models,* ed. M. Florian, North-Holland, New York.

Lowry, I. S., 1964, *A Model of Metropolis*, Santa Monica, CA: Rand Corporation.

Lundqvist, L. 1989, 'A model system for strategic metropolitan energy studies', in *Spatial Energy Analysis*, ed. L. Lundqvist, L. G. Mattsson and E. A. Eriksson, Avebury, Aldershot, pp. 245-270.

Mackett, R. L. 1983, *The Leeds Integrated Land Use–Transport (LILT) Model*, Transportation Road Research Laboratory, Supplementary Report 805, Crowthorne, Berkshire, England.

Mackett, R. L. 1985, 'Integrated land use-transport models', *Transportation Reviews*, No. 5, pp. 325-343.

Mackett, R. L. 1990a, 'The systematic application of the LILT model to Dortmund, Leeds and Tokyo', *Transportation Reviews*, No. 10, pp. 323-338.

Mackett, R. L. 1990b, 'Comparative analysis of modelling land use–transport interaction at the micro and macro levels', *Environment and Planning,* No. *22A*, pp. 459-475.

Mackett, R. L. 1990c, *MASTER Model* (*Micro-Analytical Simulation of Transport, Employment and Residence*), Report SR 237, Transport and Road Research Laboratory, Crowthorne, England.

Mackett, R. L. 1991a, 'A Model-based analysis of land use and transport policies for Tokyo', *Transportation Reviews*, No. 11, pp. 1-18.

Mackett, R. L. 1991b, 'LILT and MEPLAN: A comparative analysis of land use and transport policies for Leeds', *Transportation Reviews*, No. 11, pp. 131-154.

Mayberry, J, P. 1970, 'Structural requirements for abstract-mode models of passenger transportation" in *The Demand for Travel: Theory and Measurement*, ed. R. E. Quant, Lexington, MA: Heath Lexington Books.

McFadden, D. 1973, 'Conditional logit analysis of qualitative choice behaviour', in *Frontiers in Econometrics*, ed. P. Zaremba, Academic Press, New York.

Nakamura, H., Hiyashi, Y. and Miyamoto, K. 1983, 'Land use–transportation analysis system for metropolitan area', *Transportation Research Record*, No. 931, pp. 11-20.

Oppenheim, N. 1995, *Urban Travel Demand Modelling*, John Wiley (USA), New York.

Ortuzar, J. de D. and Willumsen, L. G. 1994, *Modelling Transport*, John Wiley & Sons, New York.

Pendyala, R. M., Kitamura, R., Chen, C. and Pas, E. I. 1997, 'An activity-based microsimulation analysis of transportation control measures', *Transport Policy*, Vol. 4, No. 3, pp. 183-192.

Prastacos, P. 1986a, 'An integrated land use–transportation model for the San Francisco region: 1. Design and mathematical structure', *Environment and Planning*, No. 18A(3), pp. 307-322.

Prastacos, P. 1986b, 'An integrated land use–transportation model for the San Francisco region: 2. Empirical estimation and results', *Environment and Planning*, No. 18A(4), pp. 511-528.

Putman, S. H. 1973, *The Interrelationships of Transport Development and Land Development*, University of Pennsylvania, Philadelphia: Department of City and Regional Planning.

Putman, S. H. 1975a, 'Urban land use and transport models: a state of the art summary', *Transportation Research*, No. 9, pp. 187-202.

Putman, S. H. 1975b, *Further Results From, and Prospects for Future Research With, the Integrated Transportation and Land Use Package (ITLUP)*, paper presented at the Annual Conference of the Southern Regional Science Association, Atlanta, Georgia.

Putman, S. H. 1975c, *Calibrating a Disaggregated Residential Allocation Model — DRAM: the Residential Sub-model of the Integrated Transportation and Land Use Package (ITLUP)*, paper presented at the Regional Science Association, British Section, 8th Annual Conference, London.

Putman, S. H. 1983, *Integrated Urban Models: Policy Analysis of Transportation and Land Use*, Pion, London.

Putman, S. H. 1991, *Integrated Urban Model 2: New Research and Applications of Optimisation and Dynamics*, Pion, London.

Putman, S. H. 1996, 'Extending DRAM model: theory–practice nexus', *Transport Research Record*, No. 1552, pp. 112-119.

SELNEC Transport Study 1971, *The Mathematical Model*, *Technical Working Paper*, No. 5, Town Hall, Manchester.

SELNEC Transport Study 1972, *Calibration of the Mathematical Model*, *Technical Working paper*, No. 7, Town Hall, Manchester.

Sharpe, R. 1978, 'The effect of urban form on transport energy patterns', *Urban Ecology*, Vol. 3, pp. 125-135.

Sharpe, R. 1980, 'Improving energy efficiency in community land use–transportation systems', *Environment and Planning*, No. 12A, pp. 203-216.

Sharpe, R. 1982, 'Local government conservation using computer modelling', *Journal of the Royal Australian Planning Institute*, No. 20, pp. 118-120.

Small, K. A. 1992, *Urban Transportation Economics*, Harwood Academic Publishers, Paris.

Small, K. A. and Rosen, S. H. 1981, 'Applied welfare economics with discrete choice models', *Econometrica*, Vol 49, No. 1, pp. 105-130.

Southworth, F. 1995, *A Technical Review of Urban Land Use–Transportation Models as Tools for Evaluating Vehicle Travel Reduction Strategies*, Centre for Transportation Analysis, http://www.bts.gov/NTL/DOCS/ornl.html.

Stopher, P. and Meyburg A. H. 1975, *Urban Transportation Modelling and Planning*, Lexington Books, London.

Wardrop, G. J. 1952, 'Some theoretical aspects of road traffic research', *Proceedings of the Institute of Civil Engineers*, Part II, Vol. 1, pp. 325-78.

Watterson, W. T. 1993, 'Linked simulation of land use and transportation systems: developments and experience in the Puget Sound region', *Transportation Research*, No. 27A(3), pp. 193-206.

Webster, F. V., Bly, P. H. and Paulley, N. J. 1988, *Urban Land Use and Transport Interaction: Policies and Models — Report of the International Study Group on Land use/Transportation Interaction*, Avebury, England.

Wegener, M. 1982a, 'A multilevel economic-demographic model for the Dortmund region', *Sistemi Urbani*, Vol. 4, pp. 371-401.

Wegener, M. 1982b, 'Modelling urban decline: a multilevel economic-demographic model for the Dortmund region, *International Regional Science Review*, No.7, pp. 217-241.

Wegener, M. 1986, 'Transport network equilibrium and regional deconcentration', *Environment and Planning*, No. 18A(4), pp. 437-456.

Wegener, M. 1995, 'Current and future land use models', *Land Use Modelling Conference Proceedings*, Arlington, Texas.

Williams, H. C. W. L. 1977, 'On the formation of travel demand models and economic evaluation measures of user benefit', *Environment and Planning*, No. 9A, pp. 285-344.

Williams, H. C. W. L. and Ortuzar, J. de D. 1982, 'Behavioural theories of dispersion and the mis-specification of travel demand models', *Transport Research*, No. 16B(3), pp. 167-219.

Wilson, A. G. 1970, *Entropy in Urban and Regional Modelling*, Pion, London.

Wilson, A. G., Hawkins, A. G., Hills, G. J. and Wagon, D. J. 1969, 'Calibrating and testing the SELNEC transport model', *Regional Studies*, Vol. 3, pp. 337-50.

BUREAU OF TRANSPORT ECONOMICS

RECENT PUBLICATIONS

REPORTS

94 *Transport and Greenhouse: Costs and Options for Reducing Emissions* (1996) $54.95 (96 0691 2)
95 *General Aviation Flying in Australia* (1996) $24.95 (96 0691 2)
96 *Quality of Rail Freight Service: The Customer's Perspective* (1997) $11.95 (96 1017 0)
97 *Transport Synergies Between Eastern Indonesia and Northern Australia* (1998) $12.95 (0 642 30504 8)

INFORMATION PAPERS (Free from the BTE)

41 *BTCE Road Construction and Maintenance Price Index* (1997)

OCCASIONAL PAPERS

115 *Trading Greenhouse Emissions: Some Australian Perspectives* (1998) $19.95 (0 642 47622 5)

WORKING PAPERS (Free from the BTE)

33 *Taxes and Charges in Australian Transport: A Transmodal Overview* (1997)
34 *Roads 2020* (1997)
37 *Tradable Permits in Transport?* (1998)
39 *Urban Transport Models: A Review* (1998)

ELECTRONIC BULLETINS

Available from http://www.dot.gov.au/programs/bte/btehome.htm
Transport and Communications Indicators (Quarterly)
Waterline (Quarterly)

INFORMATION SHEETS (Free from the BTE)

10.1 *Externalities in the Transport Sector: Key Issues* (1998)
11 *Public Road Related Expenditure and Revenue in Australia* (1998)

INTERNET

http://www.dot.gov.au/programs/bte/btehome.htm

To order free publications write to BTE, GPO Box 501, Canberra ACT 2601, Australia, telephone 02 6274 6846, or email bte@dot.gov.au. Other titles are on sale from AusInfo, GPO Box 84, Canberra ACT 2601, Australia, telephone toll-free 132 447 (AusInfo catalogue number or ISBN is in brackets).