THE UNIVERSITY OF CALGARY

Parameter Estimation in Urban Models:

Theory and Application to a Land Use Transport Interaction Model of the Sacramento, California Region

by

John E. Abraham

A DISSERTATION SUBMITTED TO THE FACULTY OF GRADUATE STUDIES IN PARTIAL FULFILMENT OF THE REQUIREMENTS FOR THE DEGREE OF DOCTOR OF PHILOSOPHY

DEPARTMENT OF CIVIL ENGINEERING

CALGARY, ALBERTA

JANUARY, 2000

0-612-49478-0

Abstract

Large scale urban models are desirable for transportation planning and policy making because they simulate and thereby help understand and take into account the interactions between the transportation system and the spatial economic system. These models draw on theories from various disciplines to understand the nature of the entire urban system and how it might evolve. To manage this complexity, the models can be divided into sub-models. Thus parts of the model can be considered individually as can specific interactions between two or more parts.

It is common to use parameter values in an urban model from other models that are somewhat different than the modelling system used for forecasting. Sometimes these other models are different mathematically, but other times they are only the sub-models of the larger model, and the difference is that their parameters are estimated from observed data, where as in the final modelling system the circular interconnections mean that most data is synthesized by other sub-models.

These theoretical differences between models considered individually and models combined into a system strongly suggest that it is appropriate to ensure that the entire model system is accurate. This can be done through an overall calibration of the entire urban model.

These issues are investigated in the context of a comprehensive land use transportation interaction model of the Sacramento, California region based on the MEPLAN mathematical framework. The model was constructed from its sub-models and a final calibration of the entire system was performed by hand, to provide a working modelling system and to gain knowledge about the characteristics of the modelling system. The model was used to analyse transportation policy in the Sacramento region to gain an understanding of the strengths and weaknesses of the model and its applicability for policy analysis.

The knowledge gained in hand-calibrating and using the model guided the development and use of an automatic parameter estimation procedure (and software), that runs the model many times searching for the "best" parameters given certain assumptions. The automated search process uses a customized version of Newton's method, and hence requires a linearization of the model using numerical derivatives. The automated search process was able to improve the goodness-of-fit measure.

The weights in the goodness-of-fit measure were subject to ad-hoc adjustment based on an understanding of the purpose and use of the model. The sensitivity of the parameters to small changes in the weights can be calculated based on the numerical linearization of the model and on the convergence criteria. This sensitivity information provides guidance towards selecting weights, but it is also valuable in suggesting the causes of lack-of-fit and hence suggesting changes in the design of the modelling system. The estimation software was extended to allow this information to be explored interactively, by showing the largest changes in both targets and parameters that would result from a change in one target's weight.

Acknowledgements

The University of California at Davis provided the data for the Sacramento MEPLAN model and the funding for the initial design of the model. Dr. Robert Johnston and Caroline Rodier helped with the initial design and the use of the model for scenario analysis.

My supervisor, Dr. John Douglas Hunt, has been a great source of guidance.

Others involved at various stages include Azhar Shah "James" Khan, Melanie Choy and Paul McMillan.

This research was funded by the National Sciences and Engineering Research Council of Canada.

To Debbie, Seán and Stephen

Table of Contents

List of Tables

List of Figures

Part 1.

Introduction, Background and Hypothesis

1 Introduction

1.1 Background

Models representing the interaction between land use and transportation have been developed to attempt to join together the tasks of urban planning and transportation planning. Of late there has been a growing understanding that the transportation system both influences and is influenced by other aspects of urban form. There has been a recognition that transportation planners must take into account how the decisions they are charged with making will influence the entire urban economy. They must also understand how changes in the urban system will influence the performance of the transportation system. In fact, in the United States recent changes in legislation make it a requirement for many transportation planning agencies to model land use effects.

Various theoretical frameworks have been proposed for representing the interdependencies between urban form and transportation systems. Many of these have been converted into operational models in academic settings. Only a few have been applied in practice to better inform policy-makers. (Wegener, 1994; Southworth, 1995). The MEPLAN framework is a modelling system that is theoretically appealing in many ways and has a documented history of over 25 years of practical application. Standard software is available for implementing MEPLAN models (ME&P, 1994).

The apparent (and legislated) desire for land-use/transport interaction models in North America could have led to a proliferation of such models in general and of MEPLAN in particular. However this has not yet happened. Published (Southworth, 1995; Hunt, 1994; EPA, 1998) and anecdotal evidence suggests that three of the main reasons why MEPLAN models are not common in North America are:

- There has been no completed MEPLAN implementation in North America, leading to suspicion that MEPLAN is somehow inappropriate for the types of policies and data in North America,
- The development of land-use/transport interaction models is expensive and time consuming, in part because of the need for a labour intensive calibration phase, and
- The standard software that is used to implement MEPLAN was dated, in that it had no user interface beyond a text editor and various command line programs when this research was

begun. (A new version of MEPLAN is currently available which may address this concern. The new version works with the MAPINFO mapping software, providing a graphical interface to spatial and network information.)

This research addresses the first two points and begins to address the third point. A complete land-use/transport interaction model of the Sacramento, California region was developed, calibrated and applied for policy analysis. The data used to develop the model was typically available North American data, and the policies analysed matched current North American concerns and used typically available policy instruments. The model used the MEPLAN framework.

The model was initially hand calibrated, but the hand calibration was "algorithmic" in the sense that numerical measures of goodness of fit were calculated and tracked for different targets that the model should reproduce. Weights were be assigned to each of these goodness-of-fit measures, allowing them to be combined into one overall goodness-of-fit measure. The calibration minimized this overall goodness of fit using optimization methods. Optimization methods were investigated "by hand" to ascertain their appropriateness.

This provided the foundation needed to investigate and automate the calibration process. This is described below and is the main thrust of this dissertation. However, throughout the research the practical needs were also considered. Thus the research project provides several extra benefits:

- A parameter estimation program has been written that can run models (MEPLAN models in particular). The program was written in Java, which is a platform independent computer language with built in support for modern user interfaces and for network computing. This is the beginnings of a modern "wrapper" to embrace and extend the MEPLAN software package.
- The MEPLAN model has been shown to be a useful tool for policy analysis in North America.
- Insight has been gained into the likely future for the Sacramento region under different policies.

1 .2 Hypothesis

The guiding idea behind the work was:

A systematic, automated parameter estimation process will provide a direct way of finding the unique parameter values that lead to the "best" model given a set of ***assumptions****, and the information from the estimation process (and the ability to keep track and investigate that information) will provide guidance towards challenging and adapting the* ***assumptions*** *in the overall goal of finding a better model.*

The word *assumptions* in this context refers to many things. Basically they can be reduced to three types:

1) the form of the model, including the way the entire urban system is abstracted into simpler parts for modelling and the form of the relationships between these parts, with the relationships expressed as mathematical equations;

2) the understanding of the errors in the model, which include the error introduced by the abstractions and simplifications as well as the error in the measured data; and

3) the understanding of the ultimate use of the model, which identifies the errors that are important for the modelling excercise and distinguishes these important errors from the unimportant errors.

1 .3 Overview of this Dissertation.

To test this hypothesis, it was first necessary to construct a state-of-the-art urban model. Chapter 2 describes urban modelling in general and provides a literature review of the state-of-the-art in urban modelling. Part 2 (chapters 3 to 7) describes the MEPLAN modelling framework, the development of the MEPLAN model of Sacramento, California, and the use of the resulting model in analysing policy. Part 3 (chapters 8 and 9) describes various strategies that can be used to estimate the parameters of urban models, and describes the approach taken with the Sacramento model, contrasting the "hand calibration" that was initially used to build the model with the "automatic calibration" that was developed to test the hypothesis. The usefulness of the automatic calibration software in adapting the assumptions is shown. Part 4 (chapter 10) offers conclusions.

2 Overviews of urban modelling

There have been several recent documents describing the state-of-the-art in urban modelling.

The Journal of the American Planning Association published a group of articles in it's winter 1994 issue. One paper by Michael Wegener (Wegener, 1994) describes twelve different modelling frameworks that are *operational*, meaning "that they have been implemented, calibrated and used for policy analysis for at least one metropolitan region." Wegener gives a brief history of the models and the people who developed them, then compares and contrasts the modelling systems on the basis of

- Comprehensiveness
- Model Structure
- Theory
- Modelling Techniques
- Dynamics
- Data Requirements
- Calibration and Validation
- Operationality, and
- Actual and Potential Applications.

Wegener's paper covers a lot of ground, and so cannot give a detailed commentary on MEPLAN, but it provides a good background and provides reassurance that the MEPLAN modelling framework is respected both theoretically and practically.

In the same issue of the Journal of the American Planning Association, Michael Batty describes the history of urban modelling in the United Kingdom and North America (Batty, 1994). His review is instructive in understanding the ebb and flow of urban modelling. Essentially, he asserts that the early models in the 1960's were too large for the computing and data collection technology of the day, and were fatally flawed by not being based on well-developed theories. This led to a partial abandonment of modelling by planners. Through the seventies and eighties the underlying theories have been more carefully developed, and the new generation of models (including MEPLAN) are more theoretically appealing yet still practical. In the meantime, planning itself has changed as a discipline, making models more important in some areas but less important overall.

One area of urban planning where large-scale urban models seem to be regaining their crucial importance is in understanding environmental impacts. The US government has been the driver of much of this. Most importantly, it legislated that new transport projects in areas where air quality standards are not attained must demonstrate that they do not have negative impacts on air quality, and that air quality forecasting must consider the interaction between transportation and land use in a consistent fashion (Intermodal Surface Transportation Efficiency Act of 1991, and the Clean Air Act of 1990). In response to this, the US Department of Energy commissioned a study of models as tools for evaluating Vehicle Travel Reduction Strategies (Southworth, 1995).

Southworth's study describes many of the same models as Wegener's study, but in more detail and from a North American perspective. His criticisms of current models (and hence areas for future research) include:

- the need to understand some of the more complex alternatives to specific trips, including telecommuting, alternative vehicles, changes in the structure of work, and changes in departure time choice. Activity based modelling and micro-simulation could address these concerns, but these techniques need to be implemented in models that are fundamentally different to MEPLAN.
- the need for a better underlying urban economic theory for poly centric cities. The MEPLAN framework is not mono-centric and so Southworth's criticism here is not directed at MEPLAN (although advances in poly centric urban economic theory could possibly enrich MEPLAN.)
- the need for a more theoretically pure and empirically calibrated understanding of dynamics. MEPLAN's mechanisms for representing dynamics are somewhat crude, but the historical lack of calibration to dynamic data is a far greater weakness. If the calibration process were more

automated and less labour intensive then modellers could spend more time calibrating the representation of dynamics.

- the need for a more interactive software environment to match the interactive nature of planning. The research here directly addresses this issue.
- the difficulty of calibration (for MEPLAN specifically). The research here directly addresses this issue.

The historical perspectives by Batty (1994) and Southworth (1995) emphasize the importance of fairly recent developmentd in urban economic theory. The theoretical advances have been incorporated by various degrees into many modelling frameworks, including MEPLAN. A useful overview of urban economic theory and models has been written by Anas (1987). He lists 5 types of urban economic models:

- Monocentric models,
- Non-Eonomic models that have contributed to economic theory,
- Mathematical programming models,
- Econometric models, and
- Regional and Inter-regional models.

Anas does not specifically mention MEPLAN, although he mentions other models that are similar to MEPLAN in many ways. Chapter 4 compares MEPLAN to the various models and theories presented in Anas (1987).

2.1 Closely related projects

The MEPLAN modelling framework has been advanced for over 25 years. An important study of calibration issues was documented by Hunt (1994) for the MEPLAN model of Naples, Italy.

A useful summary of a range of MEPLAN models is given in Echenique et al (1990). A summary of the model's theory and a list of world-wide applications is given in Hunt and Echenique (1993).

The Sacramento, California region has been modelled using the TRANUS software package, which is similar to MEPLAN in many ways. The data and results from this study are published by Modelistica (1996).

A MEPLAN model of Edmonton, Alberta (Hunt and McMillan, 1995) has been partially developed. This study will shed light on the data and policies that would influence the design and calibration of model for cities or regions in Canada.

The future transportation policies that are being considered for the Sacramento region are documented by the Sacramento Area Council of Governments (1996).

Part 2.

The MEPLAN Model of Sacramento: Theory, Development and Use

3 An Overview of the MEPLAN Modelling Framework

3.1 Concise Overview

The MEPLAN modelling framework is the result of work on a long sequence of operational urban and regional models developed by various teams - most of them under the leadership of Marciel Echenique. MEPLAN itself is a software package produced and sold by Marciel Echenique and Partners (ME&P) in Cambridge, England.

The basis of the modelling framework is the interaction between two types of markets that exist in parallel, one concerning land (space) and the activities that occupy it, and the other concerning transport. The nature of this interaction is shown in figure 1. Behaviour in these two kinds of markets is modelled as a response to price or price-like signals that arise from market mechanisms. In the land markets, production, consumption and location decisions by activities are influenced by both money price and generalized cost (disutility) signals. In the transport markets, both mode and route selection decisions are influenced by travel disutilities that include money costs and time penalties introduced by congestion delay.

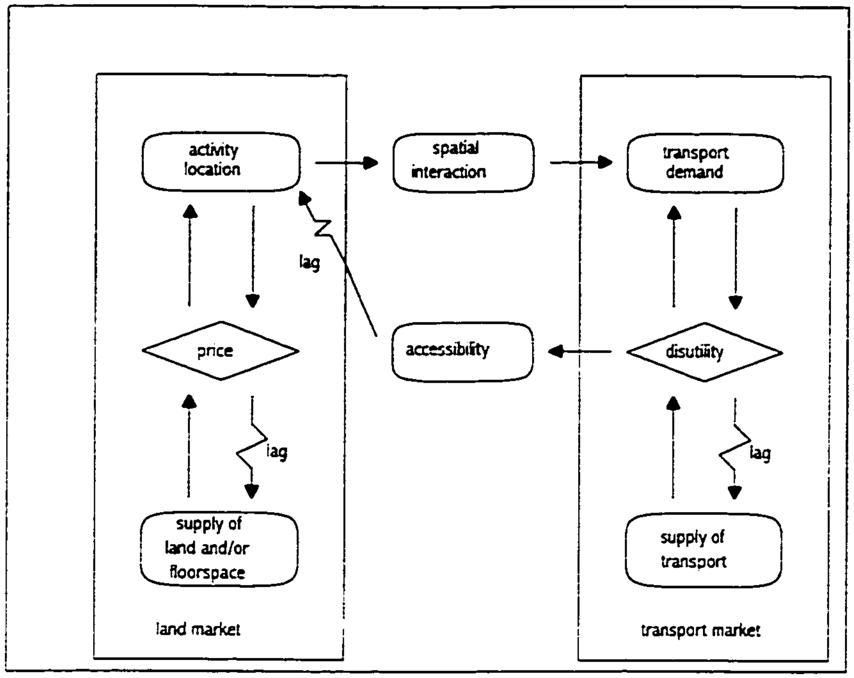


Figure 1: System of two types of markets: Markets in land and in transport and the interactions between them form the basis of the MEPLAN framework.

The cornerstone of the land market model is a spatially-disaggregated social accounting matrix (SAM) (Pyatt and Thorbecke, 1976) or input-output table (Leontief, 1941) expanded to include variable technical coefficients and the use of different categories of space, representing different types of building and/or land. Volumes of activities in the different sectors of the SAM are allocated to geographic zones using logit models (McFadden, 1974) of location choice, with the attractiveness (utility) of zones based on the costs of inputs to the producing activity (which include related transport costs), location-specific disutilities and the costs of transporting the resulting production to consuming activities.

The resulting patterns of economic interactions among activities in different zones are used to generate origin-destination matrices of different types of trips, such as home to work trips for labour movements from producers (households) to consumers (employers) and goods movements for various industry to industry interactions. These matrices are loaded to a multi-modal network representation with capacity restraint using typical nested logit model (Williams, 1977) forms for mode and route choice.

The resulting network times and costs influence transport costs, which influence the attractiveness of zones and the locations of activities, completing the 'feedback' from transport to land use.

The framework is moved through time in steps from one time period to the next, making it 'quasi-dynamic' (Meyer and Miller, 1984). In a given time period, the land market model is run first, followed by the transport market model. Then an incremental model is run concerning the changes to the next time period, and then consideration shifts to the next time period.

The transport costs arising in one period are fed into the land market model in the next period, thereby introducing lags in the location response to transport conditions.

The size and structure of the economy can be changed over time by specifying changes in the size, composition and distribution of the exogenous (exporting or Lowry (1964) 'basic') part of the economy from one time period to the next, either independently or as functions of the prices and activity levels in previous time periods. The amount and location of space are also adjusted over time in response to prices and previous activity to represent the development process. The technological coefficients can be adjusted through time to reflect changes in technology and economic structure.

The framework is fully integrated. The demands for transport are calculated directly from the interactions predicted for the spatial economic system in the land use model. This is in contrast to the much criticized process of synthesizing these demands by generating origins and destinations and then linking them using a distribution model, as is done in conventional four step transport models (with or without an associated land use model).

This framework is embodied within the MEPLAN software package, which includes various alternative mathematical forms for the different models and their components. See Hunt (1994) or Hunt and Echenique (1993) for descriptions of these mathematical forms.

3.2 The model in more detail

3.2.1 Lags and dynamics

In each market there is at any time an adjustment towards equilibrium. However, this adjustment is limited. It is limited by the impossibility of instantaneous changes in either building stock or transport infrastructure and by the imperfection of the information exchanges in the system. This leads to delays or lags in the response to price and congestion signals. The result is that the urban structure continually moves towards but almost certainly never reaches an equilibrium state. These lags are incorporated into the model representation by ordering the model operations as shown in figure 2.

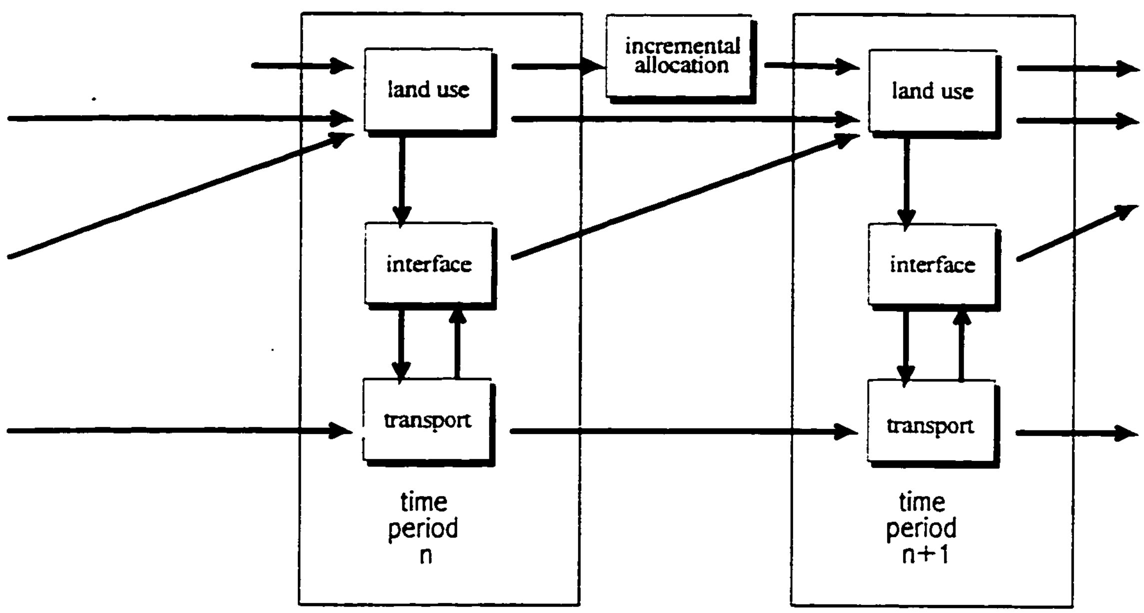


Figure 2: MEPLAN Submodels and their interactions: Temporal dynamics are simulated by ordering the sequence of interactions among the program modules at adjacent points in time.

3.2.2 Land use model

The land use model in MEPLAN is strongly based on economic theory. It begins with an input-output description (Leontief, 1941) of the economy where each economic activity draws on other activities. The input-output framework is augmented to a Social Accounting Matrix by adding in the region's households. In this way the economic model accounts for the relationships between businesses, as well as the demand for labour by businesses and the demand for goods and services by households.

The land use model is a very general system, where various components are represented by "factors". There are four standard types of factors in a normal MEPLAN implementation. The first come from an existing input-output table — factors in MEPLAN are usually chosen to represent the main industrial classifications as they are described in the input-output table. The second set of factors represents the households in the region, usually divided according to income so equity effects can be analysed and so the differences in behaviour of different income groups can be represented. The built form of the region is represented by assigning the third set of factors to the various types of building stock in the region. The effects of land use controls are represented by assigning the fourth set of factors to the different types of zoned land that may exist in a region. Beyond this, some additional factors may be specified to allow representation of specific trip types or specific travel behaviour.

It is possible to specify a functional MEPLAN model with only four factors: one representing business, a second representing households, a third representing building stock and the forth representing raw land. However, such simplicity usually fails to provide the representation needed to analyse common policies and evaluate equity concerns; so there are usually several factors allocated to each of these categories to enable an explicit representation of the differences between different types of businesses, households, buildings and land.

An input-output model such as the land use model in MEPLAN is driven by the exogenous activity. The exogenous activity is that which is "outside" of the model — it is not demanded by other components of the urban model. This includes the "basic" component of the economy — the part of each economic sector that exports from the region. It also includes the unemployed or retired households. The exogenous activity then demands or consumes other activity according to the relationships in the input-output table. The activity demanded is the generated, or endogenous, component. The basic economy and the exogenous households have certain demands that are met by attracting additional employees or households into the region, and these employees and households then attract even more activity in an infinite (but converging) loop of dependencies. For the most part, these dependencies represent actual transportation movements that can be directly represented in the transportation model.

Each economic factor (business and households) occupies buildings, and each building occupies land. The allocation of the economic factors to buildings and land gives the model its spatial component. The allocation is done according to a logit model. This approach simulates a market where a large number of decision makers choose what they consider to be the best location, each of them acting rationally within their own personal preferences and perceptions. The 'average' of all the personal preferences and perceptions is represented by a utility function, which assigns a higher utility to locations that are more likely to be chosen. The utility function includes the cost of production (which includes land costs) and the cost or disutility of transporting the factor's production. In this way the spatial allocation of activities is influenced by the disutility of travel.

3.2.2.1 Mathematical formulation of the land use model

This section describes the mathematics of the land-use submodel. It is taken largely from Hunt 1994.

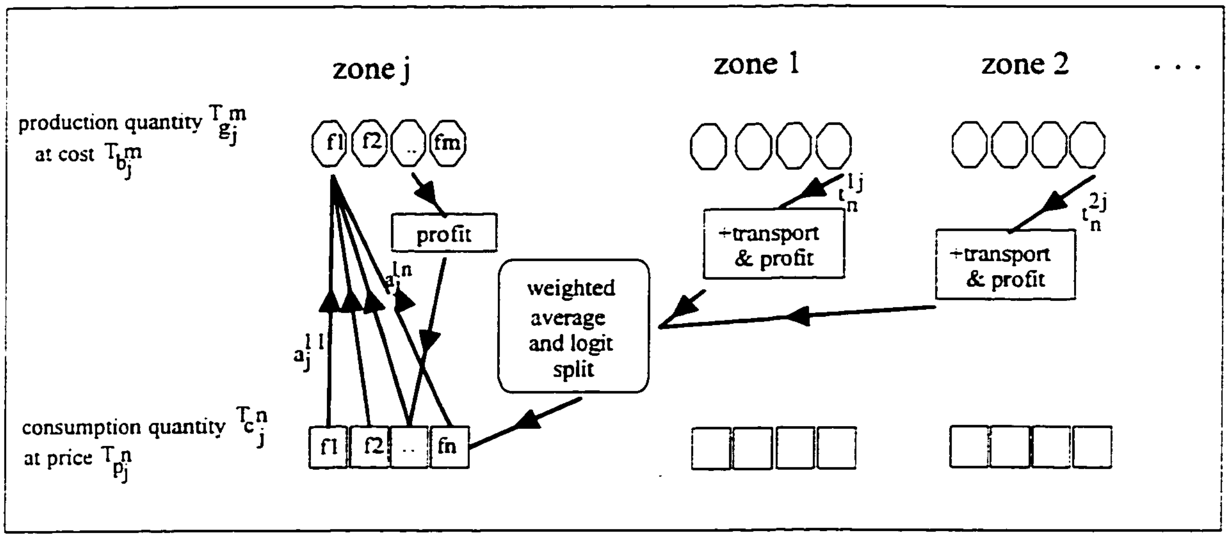


Figure 3: Relationship between production and consumption quantities, costs and prices

The demand for further production arising to satisfy consumption that is itself arising as part of the production activity within each zone is established using a modified form of input-output (Leontief, 1941) format. The general relationships between consumption, production, costs and prices in each zone are shown in figure 3. Consumption within a given zone is determined as follows:

$$ {}^{T}c_j^n = {}^{D}c_j^n + {}^{Q}c_j^n \tag{1} $$

with

$$ {}^{D}c_j^n = \sum_m a_j^{mn} \; {}^{T}g_j^m \tag{2} $$

where

j	is an index representing land-use zones;
m	is an index representing factors (categories of commodities, such as goods and services and labour);
n	is another index representing factors in the same way as m;
a_j^{mn}	is the volume of a factor n consumed in the production of a unit of factor m in zone j, called a 'technical coefficient';
${}^{T}c_j^n$	is the total volume of a factor n consumed in zone j;
${}^{Q}c_j^n$	is the exogenous component of total volume of a factor n consumed in zone j, analogous to the Lowry (1964) basic component;
${}^{D}c_j^n$	is the endogenous component of total volume of a factor n consumed in zone j;
${}^{T}g_j^m$	is the total volume of a factor m produced in zone j.

The technical coefficients, a_j^{mn}, can be fixed (as is the case in traditional input-output modelling) or variable in order to represent elasticity of demand with respect to price and/or income as follows:

For consumption that is fixed:

$$ a_j^{mn} = \mathring{a}_j^{mn} \tag{3} $$

where $\mathring{a}_j^{mn}$ is a constant parameter.

For consumption that is price elastic:

$$ a_j^{mn} = \mathring{a}_j^{mn} + \hat{a} \cdot \exp\left[-a^{mn} \cdot {}^{T}p_j^n \right] \tag{4} $$

where:

$\hat{a}$	is a constant parameter

a^{mn} is a dispersion parameter associated with the sensitivity of demand for factor *n* with respect to price

${}^{T}p_j^n$ is the price of consuming a unit of factor *n* in zone *j*.

For consumption that is both income and price elastic:

$$a_j^{mn} = \mathring{a}_j^{mn} + U^m \left[\frac{\dfrac{a^{mn}}{{}^{T}p_j^n}}{\prod_n \left[\dfrac{a^{mn}}{{}^{T}p_j^n} \right]^{a^{mn}}} \right] \qquad (5)$$

where:

U^m is the utility associated with the consumption of a bundle of quantities of various factors in the production of a unit of factor *m*

a^{mn} the parameter associated with the sensitivity of the demand for factor *n* with respect to price.

The last of these options is based on the Stone-Geary consumption function (Theil, 1980) and is used to represent the behaviour of households as utility maximizers.

The spatial element is introduced by allocating this consumption, ${}^{T}c_j^n$, among all zones according to the following formula:

$$t_{ij}^n = {}^{T}c_j^n \frac{\exp\left[\lambda^n\left({}^{T}b_i^n + d_{ij}^n + s_i^j + {}^{Q}\varepsilon_i^n + {}^{D}\varepsilon_i^n\right)\right]}{\sum_i \exp\left[\lambda^n\left({}^{T}b_i^n + d_{ij}^n + s_i^j + {}^{Q}\varepsilon_i^n + {}^{D}\varepsilon_i^n\right)\right]}, \qquad (6)$$

where

i is an index representing land-use zones;

t_{ij}^n is the volume of factor *n* produced in zone *i* and consumed in zone *j*;

d_{ij}^n is the disutility associated with transporting a unit of factor *n* from zone *i* to zone *j*; this for the person or agent making the production location decision;

s_i^j is a size term, which accounts for the *a priori* likelihood that a unit of factor *n* is produced in zone *i*;

$^{Q}\varepsilon_i^n$ is the exogenous component of zone-specific disutility associated with producing factor n in zone i;

$^{D}\varepsilon_i^n$ is the endogenous component of zone-specific disutility associated with producing factor n in zone i;

$^{T}b_i^n$ is the cost of producing a unit of factor n in zone i;

λ^n is a dispersion parameter associated with the distribution of production of factor n.

The resulting volumes of output transferred between zones, t_{ij}^n, are called 'trades'. These trades give rise to the transport demands, called 'flows', considered in the transport submodel.

The prices used in the model are endogenous. The prices and costs follow the same general relationships shown in figure 3. The price of consuming a unit of factor n in zone j, $^{T}p_j^n$, can be built up in two different ways.

One way is to treat $^{T}p_j^n$ as the weighted average cost of producing the factor in each zone plus the cost of transporting it to zone j, as follows:

$$^{T}p_j^n = \sum_i \left[\frac{t_{ij}^n}{^{T}c_j^n} \left(^{T}b_i^n + \delta_{ij}^n \right) \right] + {}^{Q}p_j^n \tag{7}$$

where:

δ_{ij}^n is the money cost for transporting a unit of factor n from zone i to zone j

$^{Q}p_j^n$ is the exogenous component of price in zone j, that can be used to help calibrate the model and to specify taxes or subsidies.

The other way is to treat $^{T}p_j^n$ as a market price that arises because of a short-run equilibration between supply and demand in the zone. In this case, an iterative process is used to establish a value for $^{T}p_j^n$ such that the demand for factor n is equal to the constrained supply in each zone. The update equation used in the iteration process can have the following form:

$$^{T}p_j^n = {}^{T}p_j^{n\prime} \cdot \left[\sum_i \frac{t_{ij}^n}{S_j^n} \right]^{\left(\frac{1}{\lambda^n}\right)} \tag{8}$$

where:

$^{T}p_j^{n\prime}$ is the price of consuming a unit of factor n in zone j in the previous iteration

S_j^n is the total available supply of factor n in zone j.

This second way is typically used to represent a short-run market equilibrium rather than a long-run equilibrium. In the short run, prices are not directly related to costs. This is used for space factors because the amount of space of a given type in each zone is fixed at a given point in time: additional space cannot be developed instantaneously nor can it be produced elsewhere and brought to the zone as a trade. An iterative process is used to establish the set of prices per unit space in all zones such that the demand for a given type of space equals the supply in each zone. This requires that the technical coefficient for this space be elastic with respect to price, thereby allowing the total demand for space in each zone to respond to the price signals and adjust towards the supply constraint. Changes in the supply of space in zones (i.e. development) occur when the entire model shifts from one time period to the next. (Within the standard MEPLAN software implementation, after equation 98 is applied the residual in equation 7 is calculated. This residual is interpreted as "profit" and is added to equation 7 so that equation 7 can be used in most of the software implementation regardless of which method is used to update prices. This profit is shown in figure 3.)

The prices established for zones in either case above determines the costs of production in zones. Specifically, the cost of producing factor m in a zone j is the result of the combination of the prices in the zone for all the inputs to the production process for factor m, as follows:

$$^{T}b_j^m = \sum_n a_j^{mn} \cdot {}^{T}p_j^n + {}^{Q}b_j^m \tag{9}$$

where:

$^{T}b_j^m$ is the cost of producing a unit of factor m in zone j

$^{Q}b_j^m$ is the exogenous component of cost of producing a unit of factor m in zone j, that, as above, can be used to help calibrate the model and to specify taxes.

The costs of production established in this way enter the volume and location decision equations described above. They also feed into the determination of prices as indicated in equation 7. This results in 'chains' of prices and costs in the submodel that run opposite to the 'chains' of demand as represented by the input-output structure. These chains of prices and costs begin wherever a market price is determined within the system because of a constraint on supply — typically for space — and they end at the prices for the factors being exported.

Running the land-use submodel involves simultaneously solving the various equations indicated above while adjusting prices in order to satisfy the supply constraints. This process uses a series of nested iterations that converge to a stable solution. Experience has shown that constraints on floorspace supply in particular play a key role by helping make this solution a unique one.

The mathematical structure of the land use model is described somewhat differently, and perhaps more simply, in the context of firm location in Chapter 7.

3.2.3 *Land use/transport interface*

The interface module converts the dependencies and relationships from the land use and economic model to actual trips for the transport model. A section of the interface module involves converting the numbers that describe the size of the economic interactions to the numbers of trips. An additional section converts the type of economic relationship to a specific trip type - or "flow". For instance the relationship between the many different types of employers and the many types of households are converted to commute trips based on the number of trips made per employee. The trip rates can be elastic to represent how changes in travel disutility will influence the number of trips made independently of the size of the economic relationship.

The interface module also allows conversion between different time scales. This allows the economic relationships to be based on a month or a year, while the transport model can represent a typical day. There is also a facility to have a different zoning system on the transport side of MEPLAN.

3.2.4 *Transport model*

The transport model takes the trips calculated from the interface module and divides them onto different travel modes and assigns them to the network. This gives indications of the level of use for each transportation link, as well as of the difficulty of travel between each zone pair for each trip type.

The transportation model represents several modes for each trip type. These are 'user-specific' modes, in that they represent the modes as perceived by the traveller. Examples of 'user-specific' modes include drive-alone, cycle, park-n-ride and transit. The entire journey consists of only one 'user-specific' mode, although each journey using these modes can use a combination of different 'states' at different points during the route. States include the use of specific vehicles, such as 'drive' and 'ride bus', but also include other transportation related activities such as 'wait', 'find parking' and 'walk'.

The transportation network is coded using a standard node and link representation. There are link types corresponding to different types of physical infrastructure, and link types corresponding to fixed-route transit service on the infrastructure. There are additional link types that work together with the states to allow specific combinations of states in certain orders. For instance, there are walk links that connect roadways and parking lots to train stations that can only be used by travellers in the 'transit walk' state, and there are walk links between parking lots and final destinations that can only be used by travellers in the 'car walk' state. This controls the availability of links to the different 'user-specific' modes so that vehicles can only be used in logical combinations. For instance, the network coding allows the use of private vehicles before the use of transit (park 'n ride) but not after.

The transport demands from the interface module are loaded onto the representation of the transport network by means of nested logit mode split and assignment models that take into account congestion effects.

The disutilities used in the mode split and assignment models are calculated from the characteristics of individual links in a very flexible manner. This allows different trip types and different transportation operators to have different preferences regarding costs, times and congestion. Similarly, the costs of travel (for feeding back to the land use model) are also built up from the characteristics of the individual links.

The zone to zone travel disutilities are the composite utility for the set of available modes, as follows:

$$h_{xy}^{k} = -\left(\tfrac{1}{\Psi^{k}}\right) \cdot \sum_{\mu} \exp\left[\Psi^{k}\left(u_{xy}^{k\mu}\right)\right] \qquad (10)$$

where:

h_{xy}^{k} is the transport disutility associated with travel for a trip of flow type k between zone x and zone y

μ is an index representing modes

Ψ^k is a dispersion parameter associated with the distribution of trips of flow type k among various modes

$u_{xy}^{k\mu}$ is the transport disutility associated with going from zone x to zone y by mode μ for trips of flow type k.

The travel money cost is the weighted average of the costs among the available modes, as follows:

$$\eta_{xy}^{k} = \sum_{\mu}\left[\eta_{xy}^{k\mu} \cdot \frac{v_{xy}^{k\mu}}{v_{xy}^{k}}\right] \tag{11}$$

where:

η_{xy}^{k} is the transport money cost for a trip of flow type k between zone x and zone y

$\eta_{xy}^{k\mu}$ is the transport money cost for a trip of flow type k between zone x and zone y by mode μ, with this built up using representative attribute values for mode μ between zone x and zone y.

v_{xy}^{k} the total number of trips of flow type k between zone x and zone y

$v_{xy}^{k\mu}$ the total number of trips of flow type k between zone x and zone y by mode μ.

3.2.5 Incremental allocation model

The incremental allocation model provides for the representation of processes that take time to occur and so are not represented in the equilibrium land use model. This includes the changes in the exogenous demand for factors and the changes in the amount of building stock or zoned land that is available. The changes can be specified at the zonal level or can be specified for the entire study area. The incremental model includes a mechanism for allocating study wide changes to individual zones. In this way the expected overall changes can be spatially allocated based on zonal attributes calculated within the model.

This allows zoning changes by governments to be responsive to the needs and demands of the market. It also allows the creation of buildings to occur in those zones where there is the greatest potential profit for developers. Exporters can choose zones where the costs of their inputs (including land) are low, and retirees can locate in zones where they have attractive land and good access to amenities.

In general, the allocation of study-wide changes to each zone is related to proportions and prices in previous time periods and to the available capacity in the zone. Various alternative forms are available for the allocation function, including linear, exponential, logarithmic and multiplicative. For example, the general multiplicative form mimics the Cobb-Douglass function (Douglass, 1934) as follows:

$$ {}_{\tau}^{\tau+1}\Delta_i^n = {}_{\tau}^{\tau+1}\Delta_I^n \cdot \frac{({}^{\tau,T}g_i^m)^{\Omega_s^{nm}} \cdot ({}^{\tau+1,T}g_i^m - {}^{\tau,T}g_i^m)^{\Omega_g^{nm}} \cdot ({}^{\tau,T}p_i^n)^{\Omega_p^n}}{\sum_i \left[({}^{\tau,T}g_i^m)^{\Omega_s^{nm}} \cdot ({}^{\tau+1,T}g_i^m - {}^{\tau,T}g_i^m)^{\Omega_g^{nm}} \cdot ({}^{\tau,T}p_i^n)^{\Omega_p^n}\right]} \quad (12) $$

where:

I	is an index representing the full set of all zones i
${}_{\tau}^{\tau+1}\Delta_i^n$	is the change in supply or exogenous demand associated with factor n in zone i from time τ to time $\tau+1$
${}_{\tau}^{\tau+1}\Delta_I^n$	is the change in supply or exogenous demand associated with factor n for the entire study area (full set of land-use zones) from time τ to time $\tau+1$
${}^{\tau,T}g_i^m$	is the total volume of factor m in zone i in time t, which acts as a size term to allow a tendency to proportional allocation (m can be the same as n in this expression)
${}^{\tau+1,T}g_i^m$	is the constraint on total production of factor m in time $\tau+1$, which together with ${}^{\tau,T}g_i^m$ acts to allow a tendency to match available capacity (again, m can be the same as n in this expression)
Ω_s^{nm}	is a parameter associated with the tendency to proportional allocation
Ω_g^{nm}	is a parameter associated with tendency to match available capacity
${}^{\tau,T}p_i^n$	is the price of consuming a unit of factor n in zone i in time τ
Ω_p^n	is a parameter associated with tendency to price-based allocation

3.3 Strengths of MEPLAN

The MEPLAN framework represents aggregate amounts of activity in an economically consistent way. The allocation of this according to random utility theory gives it a stochastic nature that is important for representing realistic spatial patterns, but the model is still an aggregate model. This level of abstraction is an important middle ground between the complexity of full disaggregate models and the unrealistic nature of non-stochastic models.

The flexibility of the MEPLAN framework in its representation of the spatial economic system, the transport system, and the relationship between these is a major strength. It allows a model to be constructed based on the final use of the model and the available data and on the economic relationships that need to be represented. The next chapter (chapter 4) describes urban economic modelling more generally and compares and evaluates MEPLAN from a perspective of urban economic modelling. The flexibility of the modelling framework should be appreciated in chapter 5, where the detailed design of the MEPLAN model of Sacramento is described.

4 A review of the MEPLAN framework from the perspective of urban economics

4.1 Introduction

Urban economics is a discipline of study concerned with the allocation of resources in an urban area, and in particular on the spatial arrangement of activities and the consumption of land and floorspace.

This chapter is a review and critique of the MEPLAN framework from an urban economics perspective. It is based primarily on "Modeling in Urban and Regional Economics" by Alex Anas (Anas, 1987). Anas does not mention MEPLAN in his review, although he mentions models that are similar to MEPLAN in various ways. The information on models of chaotic and self-organising systems is almost exclusively from Allen (1997). Other sources used for this review include Fujita (1987) and Knox (1987).

This chapter provides details on the assumptions and simplifications embodied both in the MEPLAN software framework framework and in the manner in which it is typically applied. During calibration, the information from the estimation process can provide guidance as to whether some of these assumptions and simplifications need to be challenged or adapted, by extending the modelling software or using it in a different way.

4.2 MEPLAN and the taxonomy of models

4.2.1 Monocentric models

4.2.1.1 Basic Theory

Monocentric models are models that assume a single "market" or commercial node where all goods and services are exchanged. Land outside of this single point can be used for a variety of purposes. The original formulations of the monocentric theory were agricultural in nature, and the "market" corresponded to an actual agricultural market, while the land radiating outwards from the market was allocated to the production of different types of agricultural goods (von Thunen, 1826). This established the base theory, under which the activity with the highest bid for any parcel of land would occupy that land, leading to an orderly arrangement of activities according to the various bid curves, with the bid curves related to the production function of the activity and the cost of transporting the goods to market.

In applying this agricultural theory to urban economics, the "market" is taken as the Central Business District where all labour occurs. Households are the consumers of land outside of the CBD, and their "production function" involves preferences of how much land they are willing to consume at what price. Their indifference curves for different amounts of land at different distances is related to their ability to consume "other goods and services" with the budget that is leftover after paying rent for the land and paying for the transportation necessary to commute to the CBD.

The utility functions of households are normally taken to be identical, and the willingness to pay for land at different distances is related only to different budget constraints due to different incomes. This leads to a spatial arrangement of households according to income. Under most assumptions and in most studies the wealthier households are able to purchase more land at greater distances because they have the money available to pay for larger commutes, while poorer households have little choice but to compete for small amounts of space near to their workplaces.

If a time budget as well as money budget is imposed and if some household members are not employed, then households with high salaries and few dependants will also successfully compete for the most central locations, leaving the more distant locations to households with more dependants and to households with large amounts of income from non-wage sources.

The contribution of monocentric analysis to urban economic *theory* is strongly stressed in Anas (1987).

4.2.1.2 Comparison to MEPLAN

MEPLAN adopts many of the fundamental concepts of monocentric theory. Most importantly, MEPLAN includes travel time and cost, land costs and elasticity of demand for land, and income as fundamental variables, and so can recreate the scenarios that have been investigated in monocentric theory.

The one aspect that has been included in some of the monocentric literature that cannot be directly represented in MEPLAN is a time budget. MEPLAN can include the disutility of travel in the utility of households, but this is always a linear relationship that can not represent the increasing marginal costs and diminishing marginal returns that a time budget does.

Households are typically divided into categories according to income, allowing MEPLAN to model the spatial segregation by income that occurs, but not in a continuous way.

Various actors compete with the different household categories for space in MEPLAN, but the different budgets and 'average' utility functions control which activity is more likely to outbid for space, and so the arrangement of activities and the market rent for land is still determined by a process similar to bidding process in the monocentric theory. The main difference in this respect is that although MEPLAN aggregates totals of activity, it does so using random utility theory by assuming that the aggregate activity consists of a large number of individual actors each with an individual utility function. The wide range of utility functions is expressed by a random variable, and the mean of the random variable is assigned to a function. Thus, although each actor in MEPLAN is given a single "utility function", the MEPLAN utility function is actually the mean of a very large number of utility functions. This means that at any one location type there will be certain portions of different activities that are the "highest bidders", and MEPLAN begins to simulate the actual randomness that occurs in real cities.

Perhaps the most important difference between MEPLAN and the monocentric literature is that MEPLAN *is not monocentric*. That is, no predetermined central economic location is defined in MEPLAN. In general, all types of economic activity are allowed to occur in all zones. Thus MEPLAN could at least in theory converge to a multicentric city consistent with central place theory (Christaller, 1933) or a completely mixed configuration (Anas p 37). (Note, however, that Central Place Theory is related to optimum firm size and agglomeration economies. Since MEPLAN does not include explicit representation of individual firms and firm sizes it does not embody Central Place Theory directly.)

4.2.2 Non-Economic Beginnings

This category includes some of the earlier attempts to build urban models without a strong economic background, and some of the more recent attempts to model complexity and chaos as a concept. Most of these models are very practical in that they are designed to address certain problems that needed to be analysed given certain data that was available.

4.2.2.1 Forrester dynamic model

An urban dynamic model was constructed by Forrester (1969). This model is based in control theory, and describes the dynamic relationship between variables. Different variables can respond at different rates to the levels or rates of changes of other variables. In a sense this model is a numerical simulation of a system of differential equations with time discretised. It is not an equilibrium model although with proper coefficients markets it could tend to equilibrium. The markets represented include housing markets and labour markets. Forrester's model is interesting, in that is shows how dynamic relationships can be modelled without direct recourse to economic theory. But without a strong economic theory the model needs to be very carefully calibrated to observed relationships, and Anas implies that Forrester's model was only adjusted to match Forrester's own expectations.

MEPLAN models the land and labour markets in equilibrium at any one time period. Usually land markets are "constrained", meaning that the supply is fixed in the time period and price adjusts so that demand matches supply; while in other markets price chains are adjusted so that the price of outputs (and ultimately exports) are determined from input prices. This is distinctly different from Forrester's model where market prices (and indeed prices in general) are not specifically modelled, but in any one time period the "markets" neither need to clear nor have prices based on input costs.

MEPLAN has a facility for representing dynamics, and because of this certain comparisons can be made with Forrester's model. The incremental nature of MEPLAN and specifically those processes that are modelled in the incremental model can be directly compared with Forrester's model. MEPLAN could be run in 1 year or shorter time steps, and various rate-of-change variables could be maintained in the incremental model. This would allow a more direct representation of the dynamics in an urban system than is typically done with MEPLAN. In Forrester's model the differential equations are modelled using "implicit" modelling, where variables can be functions of other variables in the new time period. In MEPLAN those variables updated by the incremental model can only really be functions of variables in the previous time period, leading to an "explicit" method of solving differential equations. The explicit method is much less stable and in general requires a careful consideration of the relationship between the scales of spatial discretization and time discretization. Theoretically, the explicit nature of MEPLAN *requires* small time steps for stability, and the 5 year time steps that are typically used may not be small enough to ensure stability.

4.2.2.1 EMPIRIC

The EMPIRIC model (Hill, 1965) is similar to Forrester's model in that it represents a numerical solution of differential equations. The main differences are:

- The coefficients in EMPIRIC were meant to be estimated using two-stage least squares estimation (or similar techniques) while Forrester's model apparently had weak, ad-hoc estimation techniques, and
- Forrester's model is aspatial while EMPIRIC uses a standard zone system, although in EMPIRIC, each zone is independent so spatial interactions are not represented.

4.2.2.1 Allen model

Peter Allen (1997) has been working on models that demonstrate the chaotic nature of cities and regions. Essentially, his models show that agglomeration economies can cause urban systems to be chaotic and self organizing. The implications of this are that history and chance are important because of extreme sensitivity to initial conditions. (This is known as "The Butterfly Effect", referring to an example from weather systems, where "a butterfly stirring the air today in Peking can transform storm systems next month in New York" [Gleick, 1987, p. 8])

Planning for chaotic systems is completely different than planning for non-chaotic systems. The future can not be predicted in chaotic systems unless the dynamic nature is controlled by muting the positive feedback in the system by adaptively adjusting inputs. Thus dynamics and history are critically important.

Allen's models use difference equations to show the different ways in which spatial systems can evolve and organize. Thus, mathematically they are similar to the Forrester and EMPIRIC models, but the emphasis is entirely different.

Allen is critical of models calibrated to cross-sectional data:

"The urban models that are used at present (see Wegener, 1994) could not generate 'alternative' cities that would have resulted from a different history. They are build by 'calibratings' (sic) them on observed behaviour for the city under study, and, therefore constitute in reality only a 'description' of the existing state of affairs. The development of a 'science of complex systems' requires the formulation of generic models, not particular ones." p23

Chaotic systems require both positive and negative feedback. Allen's models include positive feedback as agglomeration economies and negative feedback as noise and congestion externalities:

"The driving force which leads to such structure, however, is complex. Because people and jobs want to be at the centre of town, land prices tend to increase there, and at the same time, the physical conditions of noise, congestion, and pollution encountered there decrease its 'attractivity' to potential residents. These effects lead to 'crowding' effects, which tend to affect residential choice, and lead to the development of suburban areas away from the centre. After a certain time, this in turn allows retail and service activities to migrate to the periphery, and to offer jobs outside the 'centre', thus leading to a further spatial extension of the town as it grows." p43

Other positive feedback that could contribute to a chaotic system include:

- the tendency to supply new transportation infrastructure where demand is highest, and
- the increasing returns to scale in scheduled transportation services such as public transit, caused mostly by the reduction in waiting time as frequency increases.

Allen compares his models with Central Place Theory (Christaller, 1933), offering a dynamic interpretation of how central places may form and succeed or fail. Descriptions of Central Place Theory discuss the dynamic nature of how firms form and fail and how central places develop; yet the published works on Central Place Theory seldom consider the actual dynamic process (a possible exception is Viacheslav; 1997.) The long-held general acceptance of Central Place Theory suggests a general acceptance of the chaotic nature of urban systems. Thus the adoption of chaotic theories into urban planning models seems likely. Currently these theories have been adopted through non-economic models, such as Allen's or others described in Batty and Longley (1994). Eventually, these explorations into chaos theory will probably adopt economic theories in a stronger way.

The incremental model in MEPLAN could be used to represent most of the dynamic equations in Allen's simpler models. As well, the positive feedback of agglomeration economies is represented in MEPLAN. But the ability to represent dynamics in MEPLAN is weak (it is a "quasi-dynamic" model according to the definition in Meyer and Miller, 1984) and a central tenant of MEPLAN is that secondary employment is allocated using an equilibrium model, not as a dynamic process. Thus it seems that although MEPLAN could represent dynamic chaotic behaviour a proper investigation of such behaviour in urban systems should be performed using a model that combines strong economic theory with the full dynamics of a model like Forrester's, EMPIRIC or Allen's.

4.2.2.2 Lowry model

In 1964 Lowry published an urban economic model that has been extensively imitated and extended. Lowry's model is based on economic base theory, where a certain amount of employment is "exporting" or "basic", and this exporting activity drives the economy. The employees in the basic industries demand housing and other services, and other employees are necessary to fill these demands. These other employees also have needs, of course, and so an infinite but converging chain of demands are created all from the "basic" industry.

Lowry's contribution was to extend economic base theory to a spatial system divided into zones. This is done by using gravity model type formulations to allocate the secondary employment to zones based on the distance to other zones and the population in the other zones, and to allocate the population to zones based on the distance to other zones and the employment in the other zones. Lowry first examined trip distributions for shopping trips and work trips to find the coefficients for his gravity models, then simultaneously estimated all remaining parameters.

MEPLAN has many similarities to Lowry's model - in fact MEPLAN claims Lowry's model in its pedigree (Hunt and Echenique, 1993). Both are based on economic base theory, although the input-output model in MEPLAN is a comprehensive generalization of the theory to an economic treatment of production-to-consumption relationships. Both use gravity type models to allocate secondary employment based on travel difficulty to households and to allocate households based on travel difficulty to employment locations. The spatial allocation models in MEPLAN are based on logit random utility models, which makes them more behavioural and easier to interpret than other gravity models, and also allows the inclusion of price information (see below). Lowry's model does not include a direct preference for greater land consumption, instead using an upper bound constraint on density.

The other major differences between MEPLAN and a Lowry model are:

- MEPLAN is more comprehensive, with multiple types of land, floorspace, industry, employment and households; and with a choice of functions for most of the relationships. Inter-industry and inter-household dependencies are also represented.
- Travel impedences in MEPLAN's spatial allocation models are congested travel times arising from a multi-modal transportation model.
- MEPLAN's constraints on land and the elasticity of land consumption give a market clearing price for land. These prices are combined with the input-output relationships to give a price for every other factor, and all these prices are used together with the travel impedences in the allocation models. The prices can also be used to adjust the consumption and production functions that are implied in the coefficients of the input-output model. All this makes MEPLAN a much more *economic* model, while Lowry's original model is "not economic but physicalistic in nature" (Anas p. 52).
- MEPLAN includes the incremental model for adjusting the constraints on space and the arrangement of basic activity, and uses travel impedences from a previous time period, making it quasi-dynamic.
- MEPLAN allows for demand coefficients that vary with price, and for households this can be consistent with utility maximizing behaviour.

4.2.2.3 Wilson's statistical gravity model and Random Utility Theory

Wilson (1967) provided a statistical interpretation of the gravity model, showing that the logit form of trip distribution corresponded to the most likely statistical arrangement of trips given constraints on the total amount of travel. This corresponded to other developments at about the same time, most clearly enunciated by McFadden (1973), giving the random utility theory of the logit model, and corresponding measures of utility and of user benefit over a range of choices. These developments made multinomial logit, and nested logit, the formulations of choice for discrete choice modelling and provided both a statistical and economic theory for understanding and interpreting the properties and results of models.

MEPLAN uses multinomial logit and nested logit models extensively, making it consistent with the theories as developed by Wilson and McFadden. A weakness is in an incomplete use of the log-sum as the aggregate measure of the cost or utility of choosing from a range of goods. According to Random Utility Theory, when a firm or household chooses to consume from a zone that is sub optimal according to the utility functions, it is because the unique nature of that firm or household is different than the average nature expressed in the utility function, and the chosen zone is in fact optimal for that firm or household. For this reason the log-sum measure needs to be used as a measure of the costs or utility associated with a range of options — to ensure that a larger number and a wider variety of options corresponds to a lower expected cost and/or a higher expected utility.

To express this in an input-output framework would require that the efficiency of each type of firm increases when a wider range of input zones are available. Unfortunately, input-output theory typically works in terms of fixed efficiency. To get around this MEPLAN uses weighted average costs instead of log-sum costs, and a major benefit of travel (allowing more variety and choice) is lost in MEPLAN's monetary accounting (but not in its spatial allocation). This inconsistency makes user benefit and supplier benefit analysis difficult.

4.2.3 Mathematical Programming

4.2.3.1 Traffic equilibrium models

Traffic equilibrium models are included in this section of Anas (1987) because they use a non-linear constrained optimization to assign traffic to a network according to Wardrop's user-optimal conditions. Anas includes a discussion of stochastic congested assignment algorithms by Daganzo and Sheffi (1977) and Fisk (1980). Anas complains that the use of *flow rates* to describe congestion is unrealistic, since flow varies by time of day and as vehicles move through a network different links will become more congested at different times. This is an argument for explicit representation of arrival and departure times, and for modelling of vehicle movements through the network.

MEPLAN uses a stochastic congested assignment algorithm, but uses Dial's algorithm rather than the more theoretically appealing Fisk algorithm. A major advantage of Dial's algorithm is that it analyses almost all possible paths (except those that involve "backward" steps, where backward is determined from the minimum path to the final destination) in a simple algorithm that can consider all origins at the same time for a single destination. A problem with Dial's algorithm is that it is dependent on network coding, and paths through areas where shorter links are coded will be more likely to be chosen than paths involving longer links. The determination of "backward" steps is also rather arbitrary and dependent on network coding.

Anas's argument for more dynamic modelling of departure time, arrival time and vehicle movements is convincing on its own, but in a large scale urban model designed to predict the dynamics of spatial change over 30 year or more it is hard to imagine that the detailed dynamics of traffic movement on a single day are important. It may be appropriate to look at better ways of abstracting traffic characteristics beyond a single "flow rate" number for each link, but for the purpose of urban economic modelling it seems hard to believe that full dynamic assignment is necessary.

4.2.3.2 Spatial price equilibrium

Spatial price equilibrium problems were considered by Samuelson (1952), who investigated spatially separated markets. With the demand and supply of each commodity at each point being a function of the price at that point, and any price differences between two points being equal to the cost of shipping between those points (for those pairs of points between which shipping occurs), a spatial price equilibrium model will determine the shipments that will occur and the prices at each point. This can be integrated with transport models to simultaneously determine the cost of shipping.

These spatial price equilibrium problems are similar to MEPLAN in that the price of each factor at each point includes the cost of transporting that factor. The main difference is that in MEPLAN factors are not commodities but classifications of goods and services, and the great variety of goods and services means that random utility theory is appropriate. This means that some commodities will be shipped from points that are sub-optimal according to the average utility functions, because the individual commodities and businesses vary about this average in a probabilistic way. The unfortunate part of MEPLAN is that its calculation of average costs is based on trade-weighted averages that would be consistent with commodity flows, rather than the log-sum averages that are consistent with MEPLAN factors.

4.2.3.3 Network design models

Anas includes a mention of network design models, where mathematical programming approaches are used to select a subset of links from a set of feasible links to achieve certain travel cost objectives. When the problem is extended to include variation in link capacity it becomes non-convex, making the mathematical minimization difficult. This suggests that some of the more recent heuristic search algorithms (genetic algorithms or simulated annealing) would be more appropriate ways to solve for optimal network designs.

A major objective of planning should be to design networks that will perform well under a wide range of possible futures, and perform well in the short term while providing flexibility in the long term. The uncertainty of the future needs to be addressed, either directly through statistical means or indirectly by investigating a range of possible futures. This is especially the case since urban systems are likely chaotic (Allen, 1997). A more comprehensive theory of network design needs to be developed to find networks that perform well at various points in the future and are robust over a range of possible futures. A search for an "optimum" network under these conditions will be very different than the network design models described by Anas.

4.2.3.4 Herbert-Stevens model

The Herbert-Stevens (1960) model of land use allocation simulates bidding for land by maximizing the total revenue obtained by land owners. As such it can allocate activity to land in an economically efficient manner. Anas uses the Herbert-Stevens model to introduce the concept of an open city, where the utility level of the populace is set exogenously, and the population of the city will adjust so that the utility level will equal the exogenously set level. Anas also introduces the concept of land-use sequences, looking at the price of land through time and optimizing the time-discounted revenue. In other works (e.g. Fujita, 1987) both land-use sequences and the open city are examined in a monocentric context.

A direct comparison between the Herbert-Stevens model and MEPLAN is difficult, because the Herbert-Stevens model assumes deterministic utility functions and MEPLAN uses random utility theory. The constraints on land availability in the equilibrium model in MEPLAN and the elasticity of land consumption in MEPLAN are consistent with the Herbert-Stevens model, however. The open city concept is not directly comparable to MEPLAN. In MEPLAN it is possible to fix the utility level of the citizens exogenously, but the population is also fixed exogenously. This would lead to over constraint in the Herbert-Stevens model, but in MEPLAN the extra degree of freedom is in the cost of labour, which then feeds up through the model to contribute to the costs for every other good or service and ultimately defines the costs of exported (basic) goods and services.

The concept of landowners maximizing the time-discounted rent by choosing a land-use sequence can not be directly implemented in MEPLAN because MEPLAN can not have variables in previous time periods as functions of variables in later time periods. The best that can be done is to manually experiment with multiple runs of MEPLAN through time to find myopic developer behaviour functions that approximate the time-sequence optimum.

4.2.3.5 Mills linear programming model

Mills linear programming model (Mills, 1974) is based on Leontieff constant coefficient technology describing the relationships between various goods and services. The quantities of export commodities are set exogenously, and the constant coefficients on land, labour and capital correspond to the production of floorspace. An important feature is that the amount of land allocated to roads is determined endogenously for each area, by minimizing the total cost of the urban area including transport congestion. This is a normative model, in that it does not account for the fact that most congestion costs are externalities. Households in the Mills model are cost minimisers, rather than utility maximisers. Mills model centres activity around a central location where the basic goods are exported from. Hartwick and Hartwick (1974) included intermediate goods and services, and investigated the relative location of different activities. It has been extended to a multi-modal system by Kim (1979), showing that as cities get larger and central land becomes more valuable it is efficient to build transit systems and subways in central areas because they make more efficient use of land. Mills model is a linear model, with the non-linear congestion functions represented by piece wise approximations. To preserve the linear nature, the urban area needs to be described using a grid or other regular geometry, which makes the model difficult to use in practice.

The Mills model and MEPLAN are similar in that both represent various factors in an input-output framework and both take into account congested travel times. Mills model fits the entire problem into a linear mathematical programming framework, making it less realistic and more difficult to apply. MEPLAN, on the other hand, can not solve for such things as optimal land allocation to roadways and optimal time-series-use of land. One could say that the Mills model is more theoretical and lends itself to a more general exploration of concepts, while the MEPLAN model is more practical and lends itself to a specific exploration of a particular city or region.

4.2.3.6 Koopmans-Beckmann quadratic assignment

The Koopmans-Beckmann model (1957) involves assigning a certain number of plants or facilities to an identical number of sites to minimize transportation costs between the sites. The assignment is done by maximizing net rents, simulating highest bidding for land, and the question that is asked is whether this assignment represents a market equilibrium. The answer seems to depend on a number of factors, but allowing profits to vary across plants or making the problem more continuous by allowing more than one plant at a site seems to help make the optimum solution market sustainable.

MEPLAN has a more direct simulation of markets and with variable densities and elasticity of land consumption is more directly suited to modelling a large number of independent actors. The Koopmans-Beckmann model seems to have been relegated to use by operations researchers who are actually interested in assigning plants to sites to maximize profits. That is to say the use of profit maximization to represent efficient markets has fallen out of favour; direct market simulations have, for the most part, been accepted as more appropriate.

4.2.4 Econometric Models

Econometric models combine empirical data and forecasting concerns with the urban economic theory that developed in the 1960's.

4.2.4.1 The Urban Institute Model for housing market analysis

In 1975 deLeeuw and Struyk published the Urban Institute Model. The model contains four factors: households, dwellings, builders and government. Builders supply housing at a horizontal supply curve, providing any housing that is demanded above a certain cost point. Government's involvement includes housing standards, taxes and housing subsidies. Households have a utility function, with terms for housing services, other goods and services, leisure time and the wealth and racial composition of other households in their home zone. Housing services are provided from both land and capital inputs, allowing the model to include investment into housing stock by landlords directly, with landlords looking to maximize expected discounted profits. The model is run in 10 year time steps and five zones. The model is designed to match US census data sources.

The estimation of the model's parameters occurs piecemeal, with certain coefficients fixed at the beginning, then specially tailored statistical procedures are used to find the other coefficients in sequence. Anas states that "the sequential and highly selective calibration procedures ... cannot be claimed to result in any unique coefficient set, nor is it possible to claim that the estimated coefficients are such that the entire model's predictions are maximum likelihood."

MEPLAN models developers in the incremental model not the equilibrium model, so it can not simulate a horizontal supply curve, although with adequately small time steps it would be possible to approximate one. Minimal standards of housing would be difficult to implement in the equilibrium model since land owners are not directly included in MEPLAN as an economic actor, but it may be possible to use MEPLAN's constraint process in a clever design. Otherwise, minimal standards would need to be represented in the incremental model, where landowners could improve, demolish or remove from the market any properties that do not meet minimum standards. The impact of travel time on leisure time in MEPLAN is done in a linear way, which is not as appealing as the time constraint in the Urban Institute Model. The racial and wealth externalities could perhaps be modelled as non-transportable economic factors in MEPLAN, although this would be a fairly non-intuitive artificial construction. The behaviour of land owners as expected discount profit maximisers can not be done in MEPLAN as variables in previous time periods can not be made dependent on variables in future time periods.

Coefficient estimation in MEPLAN is typically done using methods similar to those in the Urban Institute Model: certain parameters are given fixed values, then other parameters are investigated and changed to find good matches to available data. So MEPLAN model calibration is, like the Urban Institute Model's calibration, quite dependent on the priorities and skills of the model builder.

Anas's criticism of this method seems harsh however, for several reasons:

- For many of the other models in his monograph Anas does not even mention any calibration, leading to the suspicion that such models are calibrated only to the model builder's expectations.
- Using prior knowledge of parameter values or parameter relationships is well accepted in parameter estimation theory (Bard, 1974; see also chapter 8).
- For complex urban models, any likelihood function will involve strong and probably unrealistic assumptions regarding the nature of the errors. Maximizing the likelihood with one set of assumptions would give different parameter values than maximizing the likelihood with a different set of assumptions. Thus the search for coefficients will require judgement and intuition regardless.
- The definition of a "Good Model" requires understanding the purpose and use of the model. A strict maximum likelihood procedure will need to be modified according to this understanding. Thus any good model building exercise will use "selective calibration procedures".

The search for theory and methods for determining the best set of parameters is the major emphasis of this dissertation. Part 3 of this dissertation discusses these issues in more depth.

Both MEPLAN and the Urban Institute Model are practical to apply to real world problems - the theory has been made to fit with the data and resources available for practical policy analysis.

4.2.4.1 NBER

The National Bureau of Economic Research model (NBER) aims at being a model of the housing market with appropriate demand and supply side representations. This is a very large scale, highly disaggregated model, modelling market disequilibrium over a sequence of years. There are sub models for employment (basic and population serving) and for demographic change. Households have models of job change, the decision to move, the formation of new households and the choice of tenure (own or rent). The supply side has models for land use, the formulation of expectations, new construction and structure conversion. The NBER model has a large number of sub classifications of employment, structure, neighbourhood, housing quality, household types, life cycle, income brackets and education.

All the different types of households are allocated to dwelling sub markets by considering each household and market in turn. This lends itself to microsimulation, and indeed the 1976 model of Pittsburgh uses microsimulation. The microsimulation technique could allow the consideration of details about individual houses, such as age, number of rooms or size, but instead these have been rolled into the alternative specific constant for the dwelling sub market. An inconsistency that would make the model's results difficult to evaluate is that the choice of dwelling sub market is done using logit techniques, but the assignment to individual dwellings is done using a total cost minimization mathematical programming technique. Each of the sub-markets do not clear in equilibrium, but the shadow rents from the cost minimization are used to update the rents in the next time year, so presumably over time the markets would tend to equilibrium.

Some may consider the disequilibrium in housing markets to be realistic, but without a strong economic theory about how urban markets clear in disequilibrium it is difficult to interpret and understand the NBER model. The general complexity of the model only makes matters worse. It would seem that in a model this complex it is even more important to conform to well understood theories, just to allow a greater understanding of the behaviour of the model. Anas does not discuss how the model is calibrated, but hints that individual relationships are calibrated to available data, and that there is no final step calibrating the entire model.

MEPLAN has coarser disaggregations than the NBER model; only a handful of different household types and structure types are normally considered. MEPLAN assumes that prices adjust in any one time period so that the market for housing does clear, with households in aggregate consuming more or less housing as they respond to price. The choice of housing sub market can be included in the variable consumption functions in MEPLAN, and the costs of the different housing markets all influence the cost of locating in each zone, making the location assignment and sub market assignment consistent in MEPLAN (although not entirely consistent with random utility theory, since MEPLAN uses weighted averages rather than log-sum averages in it's cost calculations. See section 4.2.2.3).

MEPLAN deals with aggregate quantities of goods, relying on the law of large numbers (where shares at the aggregate level are equal to the probabilities at the disaggregate level) to allow its use of logit models. Extensive disaggregation at the level of the NBER model (or down to a microsimulation level) would violate this assumption and overwhelm the current MEPLAN framework.

MEPLAN is calibrated in a similar way to the NBER model's calibration, where certain individual relationships are first calibrated or certain parameters are fixed and then the whole model is run. In MEPLAN, however, there is usually a final comparison of model outputs to known quantities, and parameters are further adjusted to achieve a reasonable goodness of fit for the entire model. During this calibration process it sometimes happens that various problems of goodness of fit lead to changes in "fixed" model parameters or even in model structures. This is desirable in that a typical MEPLAN model is designed around the characteristics of a region, allowing it to represent the necessary relationships for the policies that are to be analysed within the constraints of the data that is available. Most other models, including the NBER model, are much more fixed in their representations. The disadvantages of using this flexibility include a long and labour intensive implementation.

The flexibility available in the design could in theory allow a comparison of a wide range of model designs, each of which reflect the characteristics of unique cities, unique policies, unique data and unique analysts. In practice, however, time and budget considerations usually only allow for one design evolving over time as described above.

4.2.4.1 CATLAS - The Chicago Area Transportation-Land Use Analysis System

The Chicago Area Transportation-Land Use Analysis System (CATLAS) was developed by Alex Anas, so there is a certain bias towards it in his book. It was designed to evaluate the impacts of transportation changes on housing values, vacancies, construction and demolition. The model runs in one year periods. A nested logit model of home and commuter mode choice is used to allocate households to zones from their employment location, where the amount of employment in each zone is exogenously specified. CATLAS has a fairly detailed model of the housing industry, with logit sub models describing whether a dwelling will be kept vacant, whether a dwelling will be demolished and whether a new dwelling will be constructed. These sub models use expected rents into the future, with expected rents being simple time discounts of the current rents. In one time period, the demand for housing in a zone is a function of rent and the amount of vacant housing in the zone is also a function of rent, so supply and demand are matched in a market clearing model by finding the equilibrium price.

Anas (1987) does not specify how the travel disutilities in CATLAS are determined. They are probably exogenously specified, meaning that CATLAS can not directly model how congestion in the transportation system interacts with land use decisions.

MEPLAN is similar to CATLAS in many ways. Logit models of home location and mode choice are used. (Although MEPLAN's use of weighted averages instead of log-sums for connecting these two models means that the joint choice of home location and mode choice is *not* a nested logit model.) The construction, demolition and vacancy sub models in CATLAS could be duplicated in MEPLAN's incremental model, but in most MEPLAN implementations the housing industry is not modelled in quite so much detail. MEPLAN typically varies the density of occupancy with price to clear the overall market for space, rather than explicitly modelling vacancy.

Implementations of MEPLAN typically include greater detail in terms of different industries, different economic relationships, different types of buildings and different trip types than CATLAS. In CATLAS workplaces are exogenously given, while in MEPLAN secondary employment is allocated simultaneously with residences, and the incremental model can allocate primary (basic) employment according to model outputs from the previous time period.

CATLAS has more temporal detail than a typical MEPLAN implementation, since CATLAS is run in 1 year time steps and MEPLAN usually uses 5 year time steps.

4.2.4.1 Models of regulated European housing markets

The price regulations of European housing markets have been explored somewhat, with Anas and Cho (1986) creating a model similar to the CATLAS model. This requires sub models of illegal subleasing, and of rationing of over demanded supply by landlords or public agencies. The rationing process includes modelling of waiting lists.

MEPLAN does not have the facility to directly model such non-market behaviour as appears to occur in European housing markets. European implementations of MEPLAN have had different sectors for public and private housing (Hunt, 1994), and constraints on price and quantity could generate zone disutility measures (in monetary units) that correspond to queuing.

4.2.4.1 HOPSIM - A supply side simulation model

Arnott (1985) has developed a model of the competitive housing market in a state of dynamic "perfect foresight" equilibrium, where "perfect foresight" means that actors in one time period have knowledge about prices and conditions in future time periods. Housing stock is described as the floor area available at various quality levels and structural densities. Households are divided into various socio-economic groups, and households choose whether to rent or own, how much space to consume and (for owners) how much to spend on maintenance.

MEPLAN implementations typically do not contain a lot of detail on own vs. rent decisions and on structure maintenance decisions. A perfect foresight model is not practical with MEPLAN since the market equilibrium in each time period is based only on inputs from the previous time period.

4.2.5 Regional and Inter regional Models

4.2.5.1 Regional and interregional input-output models

At the regional level, Anas (1987) describes a number of models based on Leontieff (1951) input-output modelling. Input-output modelling is valuable at this level because of the importance of inter-industry shipping costs in firms' higher level location decisions. Anas points out some important theoretical problems with the fixed technical coefficients in standard models, especially when industries are disaggregated according to regions.

MEPLAN uses the general input-output framework, but the use of logit models to disaggregate spatially and the variation of technical coefficients, especially for households when the utility maximising formulation is used, appear to address all of Anas's concerns.

4.2.5.2 Regional econometric models

Anas describes macro-economic models, which are typically carefully calibrated to observed data, and are designed to be consistent with data on national accounts. These models do not consider actual land use or spatial structure. Their foundation in macro rather then micro economics, the lack of detailed transportation data and the lack of spatial and land use data make them quite different from MEPLAN.

4.2.5.3 Regional intra-metropolitan models

In 1972, Engle *et al* proposed a model of the Boston region consisting of three sub models: a macro-economic model of regional output, employment and income distribution, a model of long term adjustments of population and capital stocks, and a model of land use allocation. An interesting feature is the inclusion of local government decisions within the model.

MEPLAN has been used in many intra-metropolitan contexts, and the use of the input-output model gives it enough industrial detail to consider the trade flows at this level. The incremental model in MEPLAN is similar to the model of long term adjustments of population and capital stocks in the Engle *et al* model. MEPLAN has nothing approaching a macro-economic model of regional totals, but these inputs to MEPLAN often come from macro-economic models of the region.

4.2.5.4 Models of less developed nations

Less developed nations have many unique characteristics. They often have high migration to urban areas where wages are higher but unemployment is also high. They often have extensive restrictions or tariffs on imports and exports, making certain industries largely exporting and others almost strictly internal. Both labour and capital are often less mobile because of social pressures and a lack of open capital markets.

A comprehensive model was developed by Kelley and Williamson (1983) with parameters and functional forms drawn from a vast literature from the field of development economics. Once the model was developed, it was compared to actual aggregate data with apparently reasonable results.

MEPLAN has been applied to less developed nations, but its emphasis on equilibrium markets makes it difficult to directly represent legislative or social constraints on mobility of capital or labour. Similarly, the modelling of unemployment rates requires an understanding of the disequilibrium that occurs due to incomplete information and future expectations, neither of which is specifically available in MEPLAN. These would have to be represented in an indirect manner in the incremental model.

4.3 Strengths and weaknesses of MEPLAN

4.3.1 Strengths of MEPLAN

MEPLAN is a flexible, general framework based on a number of well established economic theories. Its use of behavioural logit models is theoretically and practically appealing, making interpretation of results and selection of parameters easier. The use of utility maximizing formulations of household consumption make it consistent with the monocentric literature, while the use of input-output modelling makes it easy to apply to regional problems. The inclusion of many different industrial factors allows a more comprehensive model of the economy than models that only emphasize the housing industry.

The incremental model in MEPLAN allows for a wide variety of formulations describing how development and redevelopment occur through time, which seems to be a central component of many urban economic models.

MEPLAN contains a fairly sophisticated transport model, allowing it to examine detailed transportation infrastructure plans and to produce results describing the conditions of transportation networks.

4.3.2 Weaknesses of MEPLAN

The segregation of MEPLAN into an equilibrium model and an incremental model make it difficult to model certain processes. In particular, it is difficult to represent the types of disequilibrium and market failure that occur in labour or housing markets when vacancies are high, when unemployment is high or when markets are highly regulated. The equilibrium of these markets is core to MEPLAN, and although this is appealing to many economists there is some evidence that it does not reflect the reality of household behaviour (Richardson, 1971).

The complexity of the equilibrium model means that it cannot be solved together with variables from the incremental model, with the consequence that the incremental model can only use variables from the previous time period. This eliminates any modelling of "foresight" where actors in one time period make decisions based on variables in future time periods. The fact that the incremental model can not even simultaneously solve for variables in the current time period raises questions of stability that need to be addressed by choosing appropriately sized time increments.

A new approach for integrating behavioural logit models with input-output models should be explored, where the log-sum measure of the logit model is used in the input-output model. This would eliminate certain fundamental inconsistencies in MEPLAN, and allow it to represent economically the basic benefit of increased variety that follows from increased mobility. This is already implemented to a certain degree for households, but to include it consistently for firms and households without compromising the input-output structure may be difficult.

5 The Design and Development of the MEPLAN Model of Sacramento

5.1 Overview of the Model

5.1.1 General Approach in developing a MEPLAN model

The development of a specific model using the MEPLAN framework, including the design of the details of the model, requires:

1. selecting the model structure, including the zone system and the definitions of the various factors for economic activities, trips, modes and travel states; and the functional forms for the relationships among the items in these categories, and
2. establishing the most appropriate values for the parameters of the various functions. (Hunt and McMillan, 1995)

The development process inevitably requires that compromises be struck among desired functionality, theoretical sophistication and practical limitations - particularly those related to data availability and time and resource constraints.

In this context, model design and calibration are as follows:

- 'design' is specifying what must be in the model in order to address the policies to be considered; and
- 'calibration' is getting the model operating and providing an accurate simulation with the above design intact.

Both design and calibration are accomplished as the two steps of selecting the model structure and estimating the model parameters are performed.

Design occurs with selection of the model structure and with further modification of the structure as development progresses. The model must contain certain functional aspects and categorizations in order to address policy concerns and it must also appropriately represent the relevant behavioral mechanisms and associated causal-behavioral links. These requirements are identified to the extent possible at the start of the model development in an 'initial design'. Aspects of the design may be (and usually are) modified as model development progresses in response to various practical data limitations and problems with model fit.

Calibration involves the estimation of model parameters, the modification of model structure and the re-estimation of model parameters in the search for an appropriate simulation behaviour. A lack-of-fit between model and known reality may lead to the modification of model structure: the form of a function may be altered or the categorizations may be changed. Inconsistencies in the data that become apparent after starting calibration may also lead to changes in model structure. Thus, much of calibration involves exploring the lack-of-fit that remains after parameters are estimated.

Thus, the development process is iterative in nature, where the task of calibration — itself iterative in terms of both the estimation of model parameters and the modification of model structure — gives rise to the need to revisit the model design in order to ensure that model is able to perform as intended in the 'initial design'. The selection of model structure and the estimation of model parameters are finished when the design and calibration are complete. This entire process is complex and requires considerable time and resources, but leads to a model that accurately represents the region and the behaviour within it.

5.1.2 Concise overview of the Sacramento Model

A depiction of the specific structure for a MEPLAN model is shown in a "Hunt Diagram" (Hunt and Simmonds, 1993). Figure 4 shows the Hunt Diagram for the Sacramento model. The large matrix in the middle of the diagram lists the factors in the land-use submodel and describes the nature of the interactions between factors. A given row in this matrix describes the consumption needed to produce one unit of the factor, indicating which factors are consumed and whether the rate of consumption is fixed (indicated 'f') or price elastic ('e').

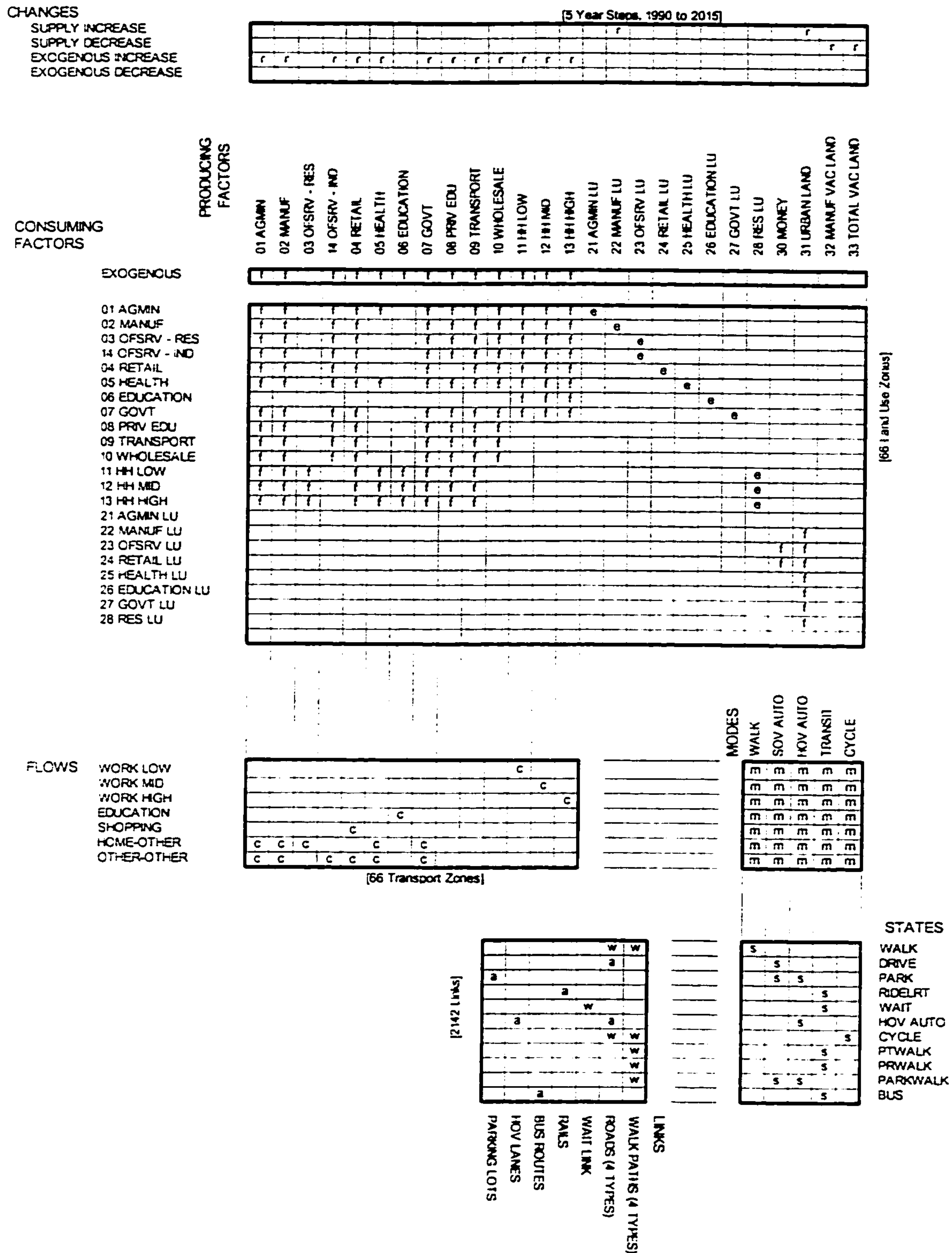


Figure 4: Structure of the Sacramento model, showing the categories and mathematical forms used in the various components. See text for symbol descriptions

The eleven industry and service factors are based on a Social Accounting Matrix aggregated to match employment location data. The resulting categories are:

- AGMIN agriculture and mining
- MANUF manufacturing
- OFSRV-RES services and office employment consumed by households
- OFSRV-IND services and office employment consumed by other industry
- RETAIL
- HEALTH
- EDUCATION primary and secondary education
- GOVT government
- PRIV EDU private education
- TRANSPORT commercial transportation
- WHOLESALE

Initially, only one factor was used to represent services and office employment. The factor was split into two during calibration in response to lack-of-fit. This is described in Chapter 9.

Households are divided into three categories:

- HH LOW households with annual income less than $20000
- HH MID households with annual income $20000-$50000
- HH HIGH households with annual income greater than $50000

These categories are based on the social accounting matrix together with the residential location data. The consumption of households by business represents the purchase and supply of labour, the consumption of business activity by households represents the purchase of goods and services by consumers.

Industries and households consume space at different rates and have different price elasticities, giving seven land use factors:

- AGMIN LU — land used for agriculture
- MANUF LU — land used for manufacturing
- OFSRV LU — land used for services and office employment
- RETAIL LU — land used for retail
- HEALTH LU — land used for health
- EDUCATION LU — land used for education
- GOVT LU — land used for government
- RES LU — land used by residences

Each of the 'c's in figure 4 represent a different space consumption function with a different elasticity. The different land use categories can be constrained in each zone, to explicitly represent different land use regulations or zoning bylaws. This makes zoning a direct input to the model - the model itself predicts how the locations of activity will change in response to zoning. Currently, constraints are placed on the amount of MANUF LU to represent zoning regulations that restrict the location of heavy industry.

There were a number of inconsistencies within the available land use data. As a result, variations in land use density with price were used to estimate the current amount of land used in each category in the base year.

Two land use factors consume MONEY. This allows differential rents to be paid for different uses of the same category of land. Currently this is done only to reflect differences in the prices for land used for different activities; the same process could be used to explicitly represent zone specific taxes and subsidies for certain activities during policy testing.

The total supply of developed land is represented by the URBAN LAND factor. This is the overall constraint in the land market - prices of URBAN LAND in each zone are adjusted until the total land consumed by all non-agricultural land uses is equal to the supply.

There is no direct representation of the amount of building stock or floorspace in the model since the relevant data were difficult to obtain. Instead, URBAN LAND represents land that has been developed and serviced. The process of developing and servicing land represents the construction of buildings and other improvements as well as various activities associated with subdividing and selling smaller parcels, including surveying, laying utilities and protecting right-of-ways for local streets.

The final two factors, MANUF VAC LAND and TOTAL VAC LAND represent the amounts of land that will be zoned and available for development in future years. Some zones have scarce TOTAL VAC LAND because certain counties have zoned much of their land as exclusively for agricultural use.

The row separated from and slightly above the largest matrix in figure 4 indicates the factors that have some of their production demanded exogenously, always according to a fixed rate. In the Sacramento model the education employees at UC Davis are demanded exogenously (a design decision made in response to lack-of-fit; see chapter 9), as are a portion of all other employment categories in each zone. The exception is for office/service employees, where the exogenous employees were all classified as OFSRV-IND employees, leaving the provision of such services to households as entirely endogenous.

The matrix at the top of figure 4 concerns the structure of the incremental allocation models. It indicates the types of characteristics taken into account in the spatial allocation (among land-use zones) of increases and decreases in the supply of space and the exogenous demands. For Sacramento, this allocation is done in response to price as well as existing distributions ('r'). The decreases in MANUF VAC LAND and TOTAL VAC LAND always match directly the increases in constraints on MANUF LU and URBAN LAND respectively, to represent the development process that converts vacant land to urbanized space.

There are two matrices just below the large matrix in figure 4. The matrix on the left indicates the structure of the interface between the land-use and transport sub models. Each row represents one of the matrices of transport demand, that is, one of the flows. It indicates the producing factors (in the corresponding columns in the matrix above) whose matrices of trades are related to that flow.

Currently, all conversions between factor and flow are constant, using values that ensure the observed total number of trips are generated for each flow given the total volumes of activity for the relevant factors.

The three matrices to the right and below this matrix concern the structure of the transport model. The matrix directly to the right shows that all modes are available to all flows ('m'). The matrix below this, on the right at the bottom of figure 4, indicates the travel states making up each mode. Each column indicates the states 's' that comprise the given mode. For example the SOV auto mode is made up of the DRIVE, PARK and PARKWALK states. The matrix to the left shows how each travel state is treated on the links in the transport network. Each row concerns a given state, indicating the links that can be used when in that state and whether there is capacity restraint operating ('a') or not ('w'). To a large extent the design of the mode split and assignment modes matches what was done in previous transport demand models of Sacramento (DKS, 1994; Modelistica, 1996).

5.2 The Design of the Model in More Detail

5.2.1 Policies and scenarios

Four transportation scenarios were specified as important before design began:

1. A "Trend" scenario with very little new transportation infrastructure,
2. A "HOV+Beltways" scenario with extensive road widenings, new highways and a considerable amount of new HOV lanes,
3. An "HOV" scenario with considerable new HOV lanes but only limited new highways and road widenings, and
4. A "Light Rail" scenario with limited expansion of the road network and an ambitious expansion of the light rail system, together with increased gasoline taxes and parking taxes

These policies meant that for the comparison project it was crucial that the model represent light rail and HOV lane usage explicitly in the mode choice model.

It was felt that the above policies would provide a very limited demonstration of the capabilities of the MEPLAN framework, and so the model was designed to be responsive to additional policies as well:

1. Changes in property tax structures, to favour specific types of land uses in specific zones,
2. Changes in zoning regulations
3. Changes in the nature of taxes on the development of new land.
4. Subsidies to encourage specific types of activity near specific types of transport infrastructure.

Thus the design of the model included a representation of different types of land and activity, and a "MONEY" factor to allow taxes and subsidies based on activity or land use.

5.2.2 *Data Considerations*

The following data were available for developing the Sacramento model:

- social accounting matrices for 1985 and 1990;
- households by income and zone for 1990;
- employment by industry and zone for 1990;
- supply of zoned land by zone for 1990;
- average prices for zoned land by category and zone for 1990;
- transportation networks by mode for 1991;
- trip volumes aggregated into various categories of purpose, mode and time of day for 1991;
- distributions of travel distances by purpose for 1991, from the SACMET travel demand model (DKS, 1994); and
- OD matrix of total trips for 1991, based on a sample of households.

The most important data limitation that was faced was a lack of information on building stock. That is the model needed to be built without knowing how much of various different types of floorspace had been developed in each zone. This led to a model structure in which floorspace was not included directly. Instead the "space" variables were somewhat more abstract, approximately representing the amount of land that had been developed.

Related to this was a lack of information on developer behaviour, i.e. what motivates developers to develop land. Data on floorspace through time would have allowed the calibration of an empirical model of developer behaviour, but without such data it was necessary to rely on certain theoretical constructs and rules-of-thumb to establish the coefficients that cntrol where development will occur and how much will occur.

The presence of a large university with a lot of students living away from home to attend the university required special consideration. Student households have different behaviour patterns and consumption patterns than other households, most notably in that they do not "work" and produce into the economy, instead choosing to consume education. This could have been modelled with the framework given adequate data about student household behaviour, but without the resources to analyse such data student households were made "exogenous" to the model. That is, many of the students at UC Davis were assigned to residences in Davis and modelled in the same way as the retired and unemployed are modelled — as net consumers choosing a location without regard to any work location. Much of the staff at UC Davis were modelled as exogenous employees, and were assigned to job locations in Davis, removing the need to directly model the mechanisms by which post secondary education activity tends to congregate in single locations. This separate representation of student households was made in response to lack-of-fit, as described in chapter 9.

The travel data was segmented in to "trip types" of home-work, shopping, education, home-to-other and other-to-other. These trips types are fairly standard and there is a wealth of collected data based on these trip types; but unfortunately these do not reflect the economic relationship that leads to a trip. For instance a trip from a work place to a doctors office could be a work-related trip if the traveller works in the health industry, but for most people this type of trip is a personal business trip. The first represents an economic relationship between the traveller's employer and the doctor, while the second represents an economic relationship between the household and the doctor. The economic relationships are what is used to spatially allocate activities in the land-use model. Given the data that was available it was necessary to make broad assumptions about the relationship between trip type and economic trades, reducing the maximum accuracy that can be obtained in modelling the effect of travel conditions on the location of activities.

5.2.3 The land use/economic model

5.2.3.1 Economic Sectors - Employment data and IO table

The division of the economy of Sacramento into factors was done according to the categories of employees and industry that were provided. There were seven categories of employees: Agriculture and mining, manufacturing, office services, retail, health, education and government. Office services was eventually divided into two types for calibration, and so this lead to the first eight factors in the MEPLAN model. These were seen to be appropriate factors because data existed on the numbers of each of these types of employees in each zone. By making the model correspond to these employees it was possible to explicitly represent the spatial allocation of different types of labour in different areas of the Sacramento region.

The input-output table that was provided did not exactly match these categories, so it was necessary to aggregate the economic categories into a set of categories that matched reasonably well with the employment categories.

Some changes and additions needed to be made to the input-output table so that all employees and households could be accounted for. The consumption of wholesale by households was seen to be shopping behaviour. In the transport model, shopping trips were to be represented as household consumption of retail. It was necessary, then, to split this household consumption of wholesale into two separate interactions: household consumption of retail and retail consumption of wholesale.

Several factors of the economy in the input-output table could not be related to the employment categories in a direct way. These were private education, transport and wholesale. It would have been theoretically possible to examine the documentation on the input-output table to determine which category or categories of employment best corresponded to these economic classifications. However, the input-output table as provided was undocumented, thus it was decided to represent these sections of the economy directly by adding them as three additional factors.

The government sector was thought to include government provided education, but the government employee category clearly did not include education employees. It was decided to leave the government category alone and represent the education employees separately. Hence education employees in the model are only endogenous by their consumption of households (representing the supply of labour by households) and by the consumption of education by households (representing the education provided to children in those households). That is to say there is no interaction in the model between education employees and the other industrial factors of the model.

This left six factors in the model corresponding to both the employment data and the input-output table: Agriculture and mining, manufacturing, office services, retail, health and government. Three factors corresponded only to the input output table: private education, transport and wholesale. One category, education, corresponded only to employment data and had no direct representation in the input-output table.

For the six factors that had a representation in both the employment data and the input output table, it was necessary to find a conversion between the numbers of employees and the monetary size of the factor. This is because the input-output table is in money units, while the employment data and the MEPLAN model work in terms of employee units. This conversion was done in the simplest way possible: by dividing the total size of the factor in terms of monetary units by the number of employees in the factor to get a production per employee number for use in all conversions.

The consumption of goods and services by households followed fairly simply from this: the input-output table had the total dollar value of this consumption for each of the industry categories, and the conversion between dollars and employees for each industry was used. The "consumption" of households by employees (that is the supply of labour by households) required additional data. The data required was the number of households, by household income, needed to supply one employee in each industry. Data was provided listing the number of households "in" each industry, and dividing these numbers by the total employees in each industry gave the coefficient that indicated the number of households needed to supply one employee.

5.2.3.2 Exogenous demand

The input-output table listed the total exports from the region in dollar amounts. The fixed conversion rates between dollars and employees was used to convert these exports into employee units for those factors that had employment data. The question then arose: in what zones are the exports produced? There was no data available to answer this question, so the "exporting employees" were allocated to all zones in proportion to the total size of that factor in each zone.

The exogenous households (the retired and unemployed) were also allocated to zones in proportion to the total number of households (by income) in each zone.

Some adjustments needed to be made however. First, there were manufacturing employees shown in zones where they were not permitted according to the land use zoning regulations. These employees needed to be moved to the zones where they would be permitted. This is caused by a slight misrepresentation of the zoning regulations. The zoning and land use categories are legal restrictions based more on occupation then on industry: there is in fact no restriction on having employees in the manufacturing industry locating on non-industrial land, there is only a restriction on employees who actually perform manufacturing in their occupation.

The second adjustment involves the University of California at Davis (as described previously under "data considerations".) There are a large number of households in Davis that consist largely of students attending the University. These students have significantly different behaviour and vocational patterns then other households, and there was significant lack-of-fit in the model when these differences were not represented (see chapter 9). However it was thought that there was not enough data to properly represent the behavioural differences of student households. Instead, some assumptions were made to calculate the number of "Davis student" households, and these households were located in Davis as exogenous households. These exogenous households had the same characteristics as the unemployed and retired households in the region: that is they did not consume any more education than any other household. For this reason, the employees at Davis also needed to be made exogenous.

5.2.3.3 Urban land

There was very little data describing the amount of building stock in the Sacramento region. This data can normally be gathered from tax assessments records and aerial photographs. However the budget in this project did not allow for any data gathering at that scale, so the decision was made to not represent building stock directly. Instead, a number of categories of "urbanised land" were created as factors.

Urbanised land represents land that has been developed and serviced. The process of urbanisation involves surveying and subdividing land, arranging to have utilities laid, protecting right of ways for local streets, selling and marketing individual parcels, and finally (and most importantly) adding buildings and other improvements. Six different factors were created to allow for the accounting of different land uses. The six factors are manufacturing land, office services land, retail land, health services land, government land and residential land.

The land consumption rates and formulas were difficult to establish. Data were available listing the maximum consumption, the average consumption, and the minimum consumption of land for five different categories (residential, commercial, services and government, health and manufacturing). Price data were also available for three different types of land (manufacturing, residential and "commercial"). This price data was said to represent the price of unimproved zoned land. For most zones the residential price was the lowest, the manufacturing price was occasionally quite a bit higher, and the commercial price was substantially higher in most zones where there was any significant amount of commercial activity.

The zoning system (described later) implied that there should be no significant difference in price between residential land and commercial land because the zoning was flexible enough that over time the two types of land were convertible into each other. Hence there was an inconsistency between the price data and the zoning assumptions. Once this inconsistency was discovered, several hypotheses could be made as to its cause. The first is that the land prices do not in fact reflect the price of raw land, but rather reflect the price of land and improvements. Under this hypothesis the premium for commercial land represents the relatively higher cost of commercial construction. The second hypothesis is that commercial uses systematically choose the best land to locate on. A third hypothesis is that the zoning system is not nearly as flexible as it was said to be, and that the premium for commercial land represents relative scarcity caused by zoning.

The zoning system was seen to be a central and important part of the model and an area where several models that were partially funded by UC Davis should be consistent with one another. Therefore, the third hypothesis was dropped from consideration and efforts proceeded to find a way of representing the first or second hypothesis.

It was decided to simply represent the premium for commercial land directly in the model. The average amount of the premium was calculated by averaging over all zones in proportion to the amount of land consumed in each zone. This left a premium for commercial of $12757 per acre over and above the cost of residential land.

The zoning system also implied that non manufacturing activities should never have to pay more for an acre of land than manufacturing activities. This is because all activities were free to locate on manufacturing land, and so there was no market reason why any activity should choose to pay more than the price of manufacturing land. For the most part the data on price reflected this - in zones where land was at all scarce (and prices were at all high) the price of manufacturing land did not drop too much below the price of residential land. It must be restated that the price of commercial land was often substantially higher than the price of residential land and manufacturing land. This premium for commercial had already been accounted for, and so the relative price of industrial and non-industrial lands seemed to reflect the zoning system given the assumptions that had already been made with respect to commercial land.

These assumptions are necessary because of the approximation arising from using a zonal "average price". Building stock and land are highly unique commodities, and the choice from among a range of parcels or leases should be represented using random utility theory. Thus a "log sum" average price would be more appropriate than an arithmetic mean average price; or a microsimulation model could represent the variation in land and buildings more directly. These options were not available within the MEPLAN framework, however.

Given these assumptions regarding price, it was possible to identify two types of zones. Most of the zones were identified as zones where there was plentiful manufacturing land and so there was no premium on manufacturing land. Seven zones were identified as zones where manufacturing land was substantially more expensive than the base land price in the zone. These zones were identified as ones where the amount of industrial land was restricted and competition for this land drove the price per acre up above the base price (the price for non-manufacturing land) in the zone. A base price was calculated for each zone, with the price for commercial land being set at $12757 higher than the base price, and the price for industrial land raising above the base price in the zones where industrial land was thought to be scarce.

The land consumption rates in MEPLAN are a function of price: the amount of land consumed for a given amount of activity will fall as price increases. The maximum consumption rate, the minimum consumption rate and the average consumption rate were given, but there was no indication of the conditions at which each of these rates would occur. The assumption was made that price was the only influence on consumption rates, and so the minimum consumption rate would occur at the location where price was highest, and the maximum consumption rate would occur at the location where price was lowest. These two points, together with the average rate and price, were used to fit the exponential consumption function used by MEPLAN.

Thus an "extra model" was created in a spreadsheet to determine the relationships between price and density for each activity type, and the parameters from this extra model were used in the larger model.

There was a constraint on the exponential consumption function however: the price elasticity of land could never rise above 1.0, otherwise when prices rise in a zone the expenditures on land (per unit of production) would drop (equation 9) leading to a reduction in average costs. Activity would be attracted by this cost reduction, leading to an *increase* in the amount of activity choosing the zone (equation 6). This would imply that land in the zone is a giffen good, which is unrealistic and inconsistent with equation 8 where prices are adjusted upwards when demand exceeds supply and downwards when supply exceeds demand. (The weakness in the model is in equation 6, which allocates production based primarily on costs rather than on utility or profits. For households, there is an advanced option in the modelling framework to allocate based on utility. A simpler solution for both households and firms is to use a version of equation 5 for the technical coefficients. The cross-elasticities of the technical coefficients implied by this equation can increase total costs when expenditures on land are less.)

For the Sacramento model, a simple approach was taken of constraining the land consumption coefficients to keep the elasticities below 1.0. Appendix A derives a rule for the relationship between certain parameter values that constrains the elasticity to be below 1.0. The resulting land consumption formulas were used in each zone to calculate the amount of land consumed by each activity at the prices that were calculated. These land numbers were taken to be the amount of urbanised land in each zone in the base year. The competition for this urbanised land would create prices that were identical to the prices that were assumed.

It should be noted that some data were available describing the "actual use" of land. Some effort was expended in attempting to use this data for the urbanized land totals. Under this configuration the model would calculate the price necessary to completely consume the land that was "actually used". It would then be theoretically possible to adjust the land consumption formulae in calibration to attempt to match the model's price data to the observed price data. However, it eventually became apparent that the "actual use data" did not represent urbanised land in actual use, because the intensity of land use from these figures was exceedingly low. These figures may have included substantial volumes of land that were essentially vacant but possibly planned for development or possibly owned by an adjacent development but not really used by that development. In any case this "actual use data" was inconsistent with the intensity of use data that was provided, and the decision was made to use the intensity data rather than the actual use data.

5.2.3.4 Zoned land

The zoning system that the Sacramento models were to emulate was exceedingly flexible. Essentially, the zoning was designed for two purposes over the long term. The first purpose was to protect certain land as agricultural land and to prohibit all development. The second was to designate certain lands as industrial lands and restrict all industrial activity to these lands. All other activities could compete with each other for the remaining land and could locate on industrial lands if necessary.

The actual zoning system in Sacramento includes a wide variety of different commercial and residential zonings at different densities. However, the distinction between commercial and residential was not seen to be important for this modelling exercise mostly because at the level of zonal aggregation it was felt that land owners in Sacramento would be successful in the long run if they wanted to convert commercial land to residential land or vice versa. Similarly, it was felt that the restrictions on density would not influence the urban system in a large way in the long term, and in any case market pressures to increase densities would lead to a rezoning of certain lands in each zone to allow higher densities.

This zoning system was represented in the model by having a factor for manufacturing vacant land and the separate factor for the total amount of vacant land. This vacant land would be converted by the incremental model into urbanised land through the development process. The factor for Urban Land represented the total amount of developed land that could be consumed by activities. An additional restriction was placed on the amount of industrial land used to allow for the increase in prices and density for industrial land where industrial land was scarce because of zoning.

5.2.4 *Interface module*

The interface module serves two major purposes. First it converts the economic relationships from the land-use model into physical flows for the transport model. Second it converts the travel costs from the transport model into monetized amounts for consideration in the spatial allocation process of the land-use model.

5.2.4.1 Trade/flow conversion

The relationships between trade relationship and physical flow are fairly straightforward and are shown in figure 4.

There are three types of work trips corresponding to commuting by the three different household income levels.

The consumption of education by households leads to school trips.

The consumption of retail by households leads to shopping trips.

The consumption of other industry (AGMIN, MANUF, OFFSRV-RES, GOVERNMENT and HEALTH) by households leads to "home-other" trips.

The consumption of other industry (AGMIN, MANUF, OFFSRV-IND, GOVERNMENT and HEALTH) by industry leads to "other-other" trips.

This provides a reasonable method of matching the trip types in the supplied trip data base with the economic relationships, but as discussed above (in "data considerations") these are broad generalisations that limit the maximum accuracy that can be obtained with the model.

The limitation of accuracy is in the calculation of costs to influence the spatial allocation of activity, and *not* in the generation of trips. Not every trip is associated with the correct trade, but every trip is still associated with some economic activity, and so the number of trips matches directly to the observed data and the distribution of trips is as accurate as would be expected from any standard synthetic method.

5.2.4.1 AM Peak factor

Four time periods were eventually modelled, the AM 3 hour peak, the PM 3 hour peak, the off peak and the AM 1 hour peak. The congested travel times from the AM 1 hour peak are what is fed back into the land use model to influence the spatial allocation of activities in the next time period. (This is largely because the model was initially designed to only represent the AM 1 hour travel conditions; and the other three periods were added in later.) In reality, of course, the accessibilities at all times of the day are important when individuals or firms make location decisions. The use of only one time period introduces inaccuracy to the degree that the range of travel conditions at different times of the day are influential in the location decision process of different factors.

5.2.5 Transport model

5.2.5.1 User Modes

The modes available to travellers in the model are Walk, Single Occupant Vehicle (SOV), High Occupancy Vehicle (HOV), Transit and Bicycle. The probability of each trip occurring by each of these modes is calculated using a single-level logit model. The utility functions describing the attractiveness of each mode to each traveller are discussed in the Model Parameters section below. They include the money cost of travel, different elements of time and mode specific constants.

This allows for the testing of various policies that might influence people's choice of how to travel, allowing the consideration of the different types of investment in different types of transportation infrastructure, including HOV and cycling facilities.

The inclusion of a wide variety of alternatives to the single occupant vehicle allows a fuller representation of the air quality effects of physical form and policy more conducive to different modes.

There are no modes for goods movement in the model. This restricts the ability of the model to predict how shipping costs influence firm location.

5.2.5.2 States

A traveller choosing to travel by a specific "user mode" can perform a number of quite different actions during the journey. Auto users need to walk to their car, then drive to a parking location, park the vehicle then walk to the final destination. These different activities that are involved in travel are represented in the model through "states".

The states are

- walk - walking the entire distance from origin to destination
- drive - driving alone for a part of their journey.
- park - parking a motorised vehicle
- RideLRT - riding the light rail system
- Wait - waiting for a transit vehicle
- HOV Auto - driving or riding in a vehicle with more than one occupant
- Cycle - cycling the entire distance from origin to destination
- PT walk - walking to a transit stop from the origin or from another transit vehicle
- PR walk - walking from a transit vehicle to the destination
- park walk - walking from a parking space to the destination

- bus - riding a bus

5.2.5.3 Link Types

These activities can occur on many different types of infrastructure, and the type of infrastructure influences the activity itself, through different speed and cost functions. Some of the link types correspond directly to physical infrastructure:

- local road
- arterial
- expressway
- HOV facility - roadways where SOV's are not allowed
- rail
- parking lot

Other link types are more abstract

- connector - an abstract link type representing the initial movement from zone centroid onto the network
- walk link - a path convenient for pedestrians or cyclists where motorised vehicles are not allowed. There are few of these in the network because most movements of pedestrians and cyclists occurs alongside roadways.
- wait link - a link representing the waiting time for a transit vehicle; travelling "along" this link represents the act of waiting.
- train service - the actual transit service that runs on the rails
- public transit walk link - a walking link convenient for transit users
- public transit final walk link - a walking link for transit users to reach their final destination.
- park walk link - a link between parking facilities and destinations
- bus service - the actual transit service that runs on the roadways

5.2.6 *Incremental Model*

5.2.6.1 Development

The incremental model drives the development of land. (Land development can not be instantaneous, so the land use model simulates a equilibrium in land constrained by the amount of land that is developed in that time period.) The amount of development that occurs over a 5 year time step is dependent on variables from the equilibrium land market at the beginning of the time period.

The model allocates development to zones according to the price at the beginning of the time period. The total amount of development in the 5-year period is specified exogenously. The amount was found by trial and error until the increase in the weighted average aggregate rent was 1% in the Trend scenario.

The development process is represented in the model by converting the factor representing excess vacant land in each zone into the factor representing developed land for that zone.

5.2.6.2 Economic growth

The economic growth throughout the time periods of the model are fixed inputs. These are simple percentage growth rates for the economy as a whole. There was no modelling of different growth rates for different sectors, and the relationships between sectors (represented by the technological coefficients) were assumed to be steady.

This is represented in the model by fixed-percentage increases in the exogenous amount of each factor. For industry this represents an increase in employees whose production is exported from the region. For households this represents an increase in the number of unemployed and retired households.

Additional scenarios could be modelled where changes in technology affect the relationships between industries, or where the economic growth is uneven between industries. Growth could also be made endogenous, so that the growth between time periods is a function of the economic performance in the previous time period.

5.2.6.3 Size terms

The incremental model also accounts for changes in the size terms between time periods. These size terms account for the fact that, all other things being equal, activities are more likely to locate in zones where more locations are available. As a zone changes in size through development its size term also changes.

(The MEPLAN software as currently available does not allow for *updates* in the size term in this way in one run of the incremental model. Instead, is necessary to perform an additional run of the incremental model when consideration of one time period is finished. This extra run subtracts out the size term for that time period; and the run of the incremental model in the next time period can then calculate the new size term, instead of *updating* the old size term.)

5.3 Model Inputs

The design of the modelling system dictates the inputs and parameters that are used. This section summarizes the fixed inputs to the modelling system. These are:

- the description of the transport network in each time period, including price information and public transit routes and frequencies ;
- the prices of land of different types in the base year;
- the amount of activity of each type in each zone in the base year;
- the demand for each factor that is exported (analogous to the Lowry basic component for each factor) in each time period, and the spatial distribution of that activity in the base year;
- the numbers of unemployed and retired households for each household type for each time period, and their spatial distribution in the base year; and
- the changes in the total amount of urbanized land in each time period.

Since these quantities are inputs, the model will not represent how changes in the urban system might influence these quantities. Hence the model as designed cannot directly predict:

- how government transportation agencies might change the transportation network in response to increased congestion or increased air pollution;
- how better transportation systems or urban environments might attract more "basic" employment to the Sacramento region; or
- how changes in prices and mobility might change the overall rate of land development.

5.4 Model Parameters

5.4.1 Types of parameters

There are many different parameters in the model that represent relationships between entities and the behaviours and characteristics of individual entities. The information to inform the values of these parameters came from many different models and from general understandings of behaviour of the different actors making up the system. The theory of how to establish the parameter values from many different models and submodels while taking into account the purpose of the model is described in Chapter 8.

5.4.1.1 Highest level parameters

Certain parameters were adjusted in the overall highest level, because they could be informed by the comparing the highest level behaviour of the modelling system to observed data, because there was a low confidence in their previous estimate, or because their values controlled key policy outputs. This overall estimation was done by running the entire model, and comparing the many outputs of the model to known values. The highest level parameters are modified to find the set that leads to a good fit between the modelled values and the known values.

5.4.1.1 Parameters from other model estimations ("fixed")

A sequential parameter estimation strategy was used, where the overall estimation of the heuristic parameters did not consider all of the available extra data or extra models. Bayesian Sequential Estimation, (where the confidence information from prior estimations is carried through into the overall estimation procedure, see chapter 8) was not used, and so parameters that could not be estimated at the highest value were fixed at their values from earlier estimations.

In other words, the final modelling system was not used to be used to establish values for these parameters for one primary reason:

- The accuracy of the parameter estimation is already better than can be achieved with the "limited view" (section 8.2.4.1) used in the overall estimation.

In some cases this is because of the data chosen for the overall estimation. Some parameters are such that inaccuracies in the parameter's value will not lead to critical inaccuracies in the model. Thus there is little reason to augment the final search process so that the highest level operation of the model can inform these parameters.

Sometimes during calibration a lack of goodness of fit can be attributed to errors in a parameter that was previously "fixed". It is not always obvious which sequential estimation process is most appropriate, and investigations of the nature of lack-of-fit can help the analyst decide whether or not to change the process. These issues are discussed further in chapters 8 and 9.

The parameters that were taken from other models or estimated using "extra data" in the lower levels of the sequential process were, and hence "fixed" at the highest level were:

- fixed technical coefficients, describing the nature of the economic interactions among factors;
- route choice preferences and sensitivity parameters;
- the number of trips per unit of trade;
- mode choice model nesting structures;
- consumption rates for space (density as a function of price) ;
- link free flow speeds and capacity restraint functions for transport networks;
- proportion of travel in the AM peak;
- values of travel time on different types of links, for both route and mode choice;
- price, inertia and land-availability parameters for allocating new exogenous production in future years; and

- price and land-availability parameters for allocating development of land between time periods.

5.4.2 *The land use/economic model*

5.4.2.1 Urban land

The land consumption rates in MEPLAN are a function of price: the amount of land consumed for a given amount of activity will fall as price increases. The maximum consumption rate, the minimum consumption rate and the average consumption rate were given, but there was no indication of the conditions at which each of these rates would occur. The assumption was made that price was the only influence on consumption rates, and so the minimum consumption rate would occur at the location where price was highest, and the maximum consumption rate would occur at the location where price was lowest. These two points, together with the average rate and price, were used to fit the exponential consumption function used by MEPLAN. The estimation and constraints on these parameters are described in section 5.2.3.3.

Thus, a separate simple model of land consumption densities at different prices was constructed, and the parameters of this model established using a simple two-point curve fitting. These parameters were not adjusted in the overall simulation because the unavailability of suitable target data to inform these parameters suggested that the values estimated in the simple submodel should not be discarded. This illustrates the strengths and weaknesses of the sequential estimation technique. The lack of data at the overall level was overcome by constructing a simpler submodel of land consumption, and estimating the parameters for the simpler submodel. But it is still possible that the overall modelling framework could have further informed these parameter values. The sequential technique forces the analyst to choose between the parameter estimates from the simpler sub models and the parameter estimates that may be established at the highest level. Chapter 8 describes Bayesian techniques and simultaneous estimation that could allow for parameter values informed by all the data and theory.

These land consumption formulas were used in each zone to calculate the amount of land consumed by each activity at the prices that were calculated. Hence the amount of land consumed in each zone is a synthetic value.

5.4.2.2 Technical coefficient calculation

The technological coefficients describing the relationships between economic factors were from the state's input-output table, adjusted for Sacramento as part of a separate modelling effort (the IMPLAN model). The input-output table had substantial detail as to various industry types by SIC code. These were aggregated to match the factors that were chosen.

As described above in section 5.2.3, some of the factors were aggregated to match employment data. These factors could then be converted into employee units from dollar units by a constant conversion factor representing the output per employee in that industry. Other factors were aggregated to match population (household) data. These factors could be converted into household units from dollar units by a constant conversion representing the income per household in each industry.

5.4.2.3 Land use parameters estimated at the highest level.

The production allocation dispersion parameters were estimated in the overall estimation process. These are the parameters that control the relative size of the unobserved term in the spatial decision making process. The unobserved term captures two types of 'error'. First, it captures the variation in the perceptions and preferences of the decision makers. Second, it captures the effect of all those variables that are not included in the model and that make each individual location unique.

Both of these contributions to the unobserved term add randomness to the system, and in a spatial allocation system randomness leads to more variation in location and hence longer trips. Because of this, the primary 'targets' used to tune the dispersion parameters are the trip length distributions from the existing SACMET model (DKS associates, 1994).

5.4.3 Interface module

5.4.3.1 Trade/flow conversion

The number of trips per unit of trade was fixed at a constant value. This constant was calculated by dividing the total number of observed trips of a given type by the total amount of trade associated with that trip type.

5.4.3.2 Peak hour representation

From observed data, it was known what portion of total trips occur in the peak hour, and what portion of those are in the "reverse" direction of returning to the generator. These were set up for the 4 time periods of the transport model:

- AM 1 hour peak
- AM 3 hour peak
- PM 3 hour peak
- Off peak (midday)

These different parameters were put into different files, and various DOS batch files were written to ensure that the correct parameters were used for the correct purposes. Generally, the AM 1 hour peak parameters are used since the congested times from the AM 1 hour peak are used as the representative travel disutilities that impact activity location in the next time period.

5.4.4 *Transport model*

5.4.4.1 Network cost of travel

The money cost of vehicle transport was set at 7 cents per mile. This value came from the TRANUS model of Sacramento.

5.4.4.2 Disutility functions

The disutility of travel by different modes was assumed to consist of a mode specific constant for each mode plus a linear function of travel time and travel cost. The values of travel time (i.e. the trade-off rate against travel cost) were taken from the Tranus model of Sacramento. For work trips, these values are $11.25/hr value of time for high income households, $5.05/hr for middle income households and $1.88/hr for low income households.

For education trips the value of time was given as $3.75/hr. For shopping and other trips the value of time was $4.69/hr.

Cycling and walking trips have the same value of time, but it is calculated directly from the distance on the network by assuming speeds of 9 mph for cycling and 3mph for walking. This made it unnecessary to directly model the speeds of non-motorised modes on roadways.

The overall attractiveness of each mode above and beyond the details of the trip are captured in the mode specific constant for each mode. The mode specific constants were adjusted in the calibration process. The TRANUS utility functions were tuned to the TRANUS model, and so are not completely appropriate for the MEPLAN model, hence these parameters needed to be tuned in the overall calibration, giving the model enough flexibility to match observed mode shares almost exactly.

5.4.4.3 Capacity restraint

The link free flow speeds were given for each link from the network provided. Standard BPR cost-flow curves were used to calculate congested travel times. For transit, the standard MEPLAN functions calculate vehicle delay for loading and unloading, and will also calculate queue delay for loading and unloading.

5.4.4.4 Link preferences

MEPLAN has a different set of utility functions for assignment. For assignment the values used are:

- $2.30 per mile for walking
- 15 cents per mile for driving
- 8.33 cents per minute
- 16.666 cents per minute for waiting
- Average HOV occupancy (2.0 for work trips, 2.5 for other trips)
- dispersion parameter for route choice.

These values are from the TRANUS model of the Sacramento region (Modelistica, 1996)

5.4.4.5 Time periods

The transport model was specifically set up to be run for each of the four periods, to allow an investigation into the all-day travel conditions and to allow emission forecasting. Practically, this involved changing the capacity of each link to a "3 hr" capacity or a "12 hr" capacity so that the volume/capacity ratios for each link would still be consistent with the formulas and parameters used to predict congested travel times.

5.4.4.6 Mode choice dispersion parameters

The utility functions are only half of the logit mode choice model. The remaining component is the dispersion parameters. These parameters control the amount of random variation there is in people's choice of mode beyond what can be explained by the utility functions. These can be estimated in the overall estimation since the OD-matrices by mode contain information about how likely people are to switch modes under a variety of conditions.

5.4.5 *Incremental model*

5.4.5.1 exogenous employment/household location

The allocation of new exogenous ("Lowry Basic") employment to zones is performed by the incremental model. This is done in inverse proportion to the 'cost of production' but in direct proportion to the existing arrangement of activity of that type.

The allocation is inversely related to costs of production, since the income of exogenous households (retired or unemployed) or the price of exported goods is determined by conditions outside of the region. With this income or price fixed the activity will search out the lowest cost.

Allocating production in proportion to the existing distribution builds a certain amount of inertia into the model. This is appropriate in the case of retired households, for instance, because new retired households may well have been working households in the previous time period. For industry the allocation according to the existing distribution reflects the fact that the same firms are likely to be producing for both the local market and the export market.

5.4.5.2 inertia/size terms

The incremental model updates the size term used in the spatial allocation of activity (see section 5.2.6.3) The size term is the log of previous total production. This is also an inertia term in that if a lot of new activity moved to a zone in a time period, that would increase the size term and make a similar movement more likely in the next time period. Theoretically, the size term should be a coefficient between 0.0 and 1.0 multiplied by the log of the size. A coefficient of less than 1.0 indicates that the larger zones are as homogeneous as the smaller zones. This is contrary to *a priori* expectations, but sometimes occurs because of the way zoning systems are constructed. Without any empirical evidence to support this, however, the coefficient was kept at 1.0.

5.4.5.3 development

The total amount of development is a fixed input, but parameters control how the development is allocated to different zones. This is done in direct proportion to the amount of vacant land in a zone. This means that if land costs were equal then an equal percentage of vacant land would be developed in each zone.

The percentages are scaled by a factor of the cost raised to the power of 1.5. More expensive land means that developing a given parcel is more likely to be profitable, as land consumers are willing to trade-off land space against floor space. This is given an exponent greater than 1.0 since at very low land prices very little land is likely to be profitably developable.

6 Using the model for Policy Analysis

6.1 Introduction

This section compares a range of different policy options simulated for the time period from 1990 to 2015. Five different scenarios are compared with a trend scenario, three representing different long term and short term government transportation supply scenarios, one representing incentives and disincentives to encourage land use patterns complementary to an improved transit system, and one combining the land incentive scenario with higher prices for private vehicle use.

The scenario predictions differ in mode choice, firm location, residential location, development patterns, Vehicle Kilometres Travelled (VKT) and land rents. Various land uses are seen to be bidding against each other for land, and the differences in spatial arrangements between scenarios are quite complex and demonstrate the richness of the model. Certain changes take time to develop, demonstrating the need for a long-term approach to planning.

This chapter shows how the model is used, which provides some indication of what makes a model 'good' model for policy analysis. An understanding of what makes a model 'good' is essential for establishing appropriate parameter values.

This chapter describes results from the "hand calibrated" model. The calibration of the model is not described until chapter 9, where it can be compared with the "automatic calibration" that was developed in response to what was learned from applying the model, as described in this chapter.

6.2 Scenario results

6.2.1 Minimal Construction (Trend Scenario)

6.2.1.1 Description of the Scenario

The base case scenario is essentially the "current plan constrained" option in the Sacog 1996 Working Paper #3 (SACOG, 1996). It was created by removing the most expensive projects from the existing plan, to fit within the most conservative estimates of how much funding can be expected over 20 years. The projects removed were light rail extensions, construction of HOV lanes on freeways, and a set of major highway improvements to Routes 70 and 99 in the Marysville/Yuba City area. These were the most costly projects in the current plan, and would require more funding than is likely to be available over the 20 years. The resulting scenario involves little transit expansion, a modest amount of HOV lane construction and a respectable amount of road building to 2005, with very little expansion beyond 2005.

6.2.1.2 Discussion of Results

The land development over the 25 years of the trend scenario occurs north, east and south of the City of Sacramento. Figures 5 and 6 are maps of the Sacramento region. Very little development occurs in Yolo county, since most of it is zoned exclusively for agriculture. The increase in activity is shown in figure 7. The increase largely occurs in the existing built up areas northeast, east and immediately south of the CBD. For the most part, activity follows land-development, although there are increases in density in many of the zones. The density increases in the model represent redevelopment as well as increased crowding in existing building stock.

The Sacramento Region

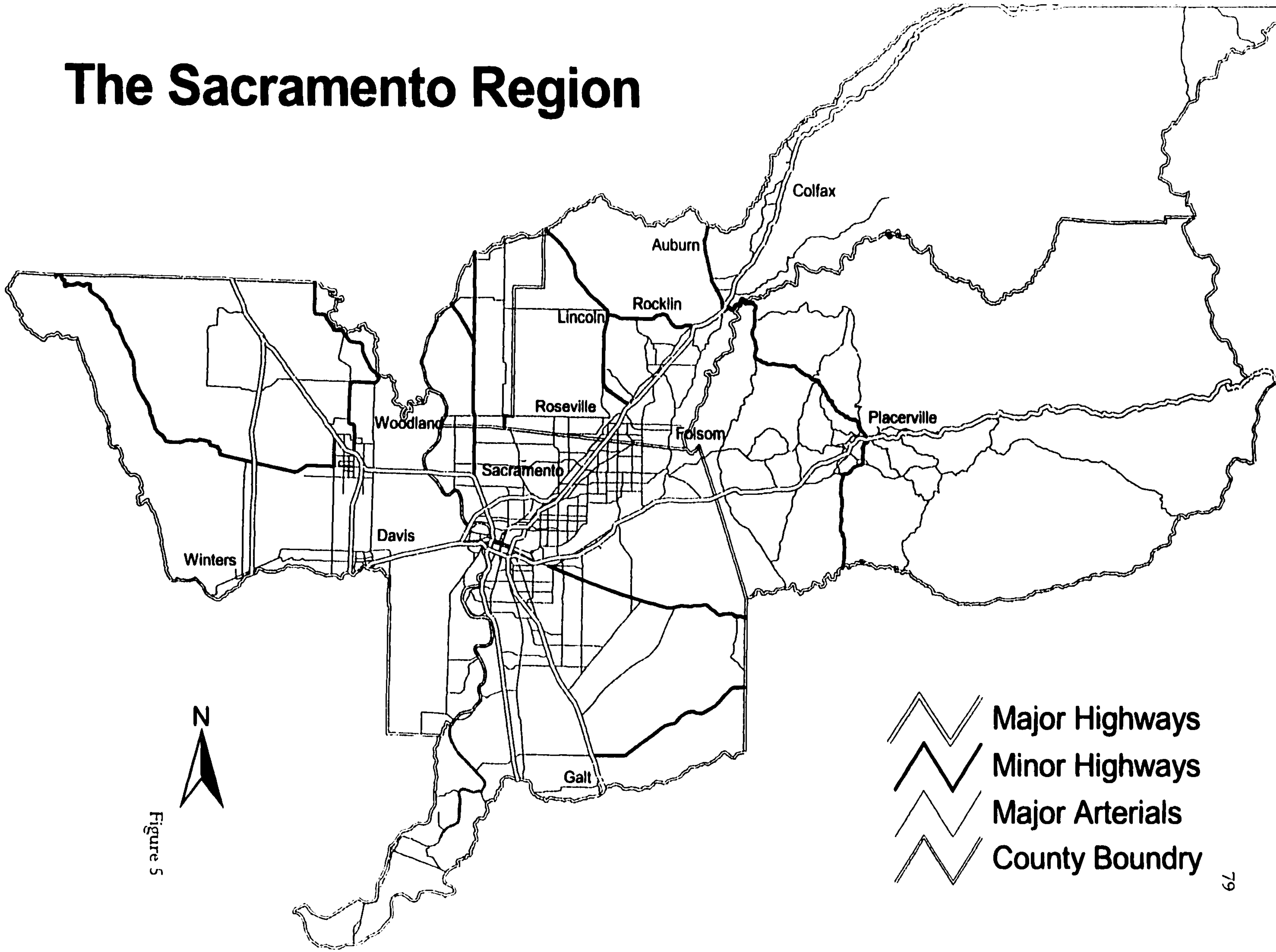


Figure 5

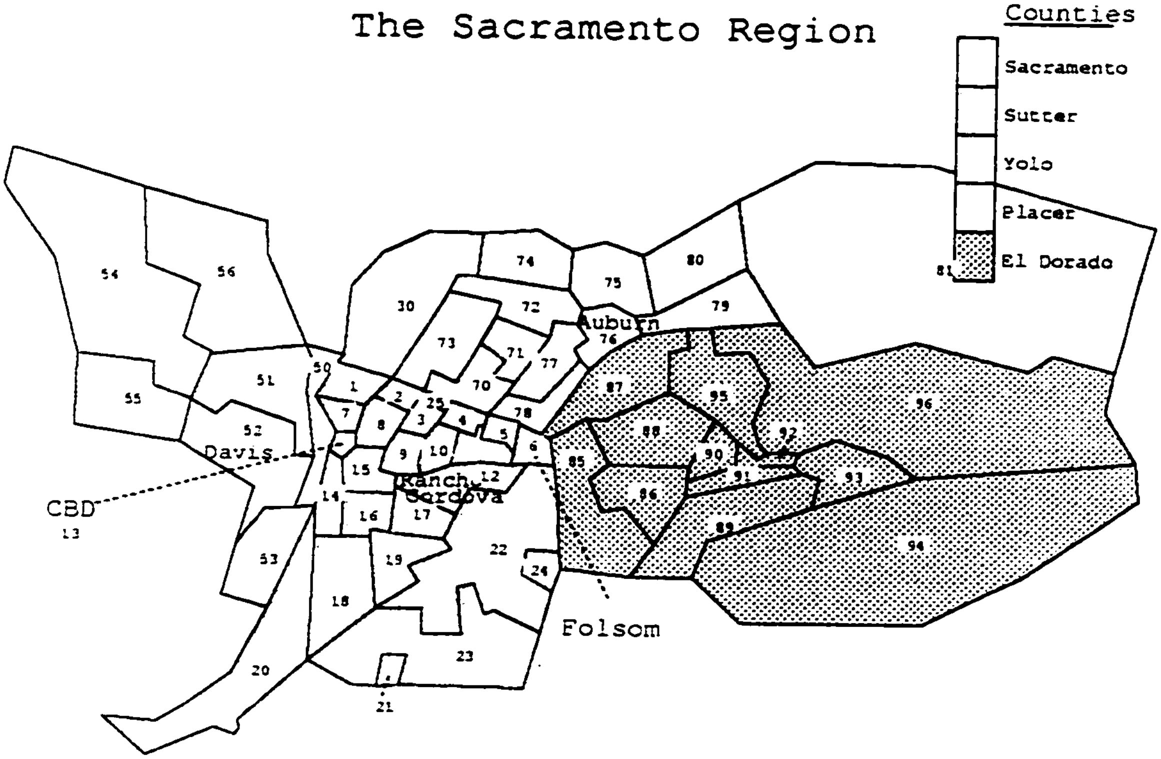


Figure 6: Map of the Sacramento region showing zone numbers, counties and selected cities

Figure 7: Maps of the allocation of increases in population and employment in the Trend scenario, 1990-2015. A dot represents 500 households in the top map and 500 employees in the bottom map

This shows two important elements of the model. Land consumption elasticities allow greater density in existing developed land when prices rise. The developer submodel allocates new land development to zones in response to price. The parameters that control these effects were not estimated using the overall modelling framework, but were fairly simply established using rules of thumb. As this is an important part of the model the values of these parameters should perhaps be confirmed or adjusted by collecting further data on land consumption and developer behaviour. Disaggregate data on land consumption would be useful, as would time series data on development.

The total amount of land development in the region was set to give very small increases in the weighted average rent. As the region grows a larger proportion of activity moves into more distant and cheaper zones, and so most zones must increase in value fairly substantially to keep the weighted average rent relatively constant. This is to be expected: as economies grow the peripheral areas become valuable and the central areas become more important. The central areas have the largest absolute rent increase, but the peripheral areas have the largest percentage increase. The areas west of Sacramento City (in Yolo county) have substantial rental increases, reflecting the scarcity of developable land in these zones.

The land markets are key to the MEPLAN modelling framework. The direct representation of land and floorspace markets and the bidding for space that occurs make MEPLAN consistent with most of classic urban economic theory. The resulting rent values are important outputs that would not be available in a less behavioural model. MEPLAN is not a micro simulation model, and so the aggregate behaviour of groups of decision makers are modelled instead of the individual decisions of random decision makers. This makes each factor into a commodity with only one price per zone, which is less realistic but also much simpler than modelling the full workings of heterogeneous markets. The price data was matched exactly in the Sacramento implementation of MEPLAN, because a simple relationship between price and density was used to establish the amount of land used in each zone. If more accurate data on developed land and floorspace amounts was available then it would have been possible to try to find the parameters that provided the best match to all of the land price and land quantity data. This could have been done in the overall calibration or in a separate, sequential estimation using a submodel. If disaggregate data were available or were collected a disaggregate "extra" model could have been used to establish the values of these parameters.

In figure 7 it can be seen that there is a large concentration of additional employment in the CBD area, while there is no increase (in fact a slight decrease although this is not visible in the figure) in households in these zones. This reflects a classic urban economic phenomena where various commercial activities can outbid residential activities. The steeper bid-rent curve of commercial activities is not 'hard-coded' into the model, but follows from the various data sources used to establish the model's parameters. These include the economic interaction data in the Social Accounting Matrix (from the IMPLAN model that was provided), observed trip length distributions by trip type (which was used in the overal estimation to establish location choice parameters), land consumption rates (which were considered in the separate land consumption model) and travel costs (which were established from the SACMET (DKS, 1994) and TRANUS (Modelistica, 1996) models.

The mode choice in the Trend scenario is predicted to change over the 25 year period as shown in table 1.

	Percentage of person trips by each mode				
Year	SOV	HOV	Transit	Cycle	Walk
1,990	47.57	35.86	3.15	4.16	9.26
1,995	47.24	36.39	2.76	4.26	9.35
2,000	46.75	36.81	2.66	4.45	9.33
2,005	46.47	37.18	2.51	4.61	9.23
2,010	46.52	37.07	2.36	4.85	9.2
2,015	46.04	37.69	2.33	5.24	8.7

Table 1: Percentage mode split in the trend scenario, 1990 to 2015

Generally, mode split is relatively constant. Transit loses mode share, with HOV and Cycle gaining mode share. There is a drop in SOV mode share, although the gain in HOV mode share is large enough such that the total mode split to private vehicle still increases. The percentage of trips made by non motorized modes increases slightly as well, although there is a shift from walking to cycling.

The changes in mode choice are small amounts, but they are nevertheless considered to be important outputs. A small change in transit ridership forecasts can strongly influence the perceptions of mobility and congestion and transit profitability. Mode choice numbers are, rightly or wrongly, considered to be essential outputs from a modelling exercise. Thus the parameter estimation process has to be such that the overall model can match observed values almost exactly.

The longer trip distances in the later years (especially for work trips) make HOV and Cycle more attractive as compared to SOV and Walk. The trip length distribution for work trips in the calibration year together with the behavioural assumptions and parameters suggested that work trip length was constrained not so much by preference but by the physical arrangement of home and workplace opportunities. That is, when the economic and physical size of the region grows people have more opportunities to live and work further apart, and more people choose to do so. For shorter trip types, such as education trips, the trip lengths in the calibration year appear to be constrained by preferences, and so a larger region does not mean longer trips. In fact, the greater future congestion in the trend scenario causes people to travel shorter distances for education to keep the travel times down.

6.2.2 *Extensive Highway Construction (Beltway Scenario)*

6.2.2.1 Description of the Scenario

The Beltway scenario involves substantial construction of new roadways and HOV lanes. This new construction is introduced by 2005, and so affects land use in 2010 and 2015.

6.2.2.2 Discussion of Results

An important shift in land use that occurs with this scenario is the increasing importance of the Rancho Cordova commercial node east of the City of Sacramento and west of Folsom. The losers of commercial activities are more distant zones, including the cities of Auburn and Folsom. These peripheral cities are a little more like 'bedroom communities' in this scenario, as the roadway expansion allows industry to locate further away from the households that it serves and employs. For instance, retail activity can shift from local commercial to 'big-box' retailing as a result of increased roadway capacity.

These movements of industry were found to be important in allowing the overall amount of travel to adjust to transportation policy and infrastructure. Chapter 7 discusses industrial location choice more fully. The factor inter-relationships in the Social Accounting Matrix control much of this movement, together with the location choice dispersion parameters estimated primarily from the trip length distributions. This scenario, and the results in chapter 7, show the importance of .

inter-industry migration and suggest that it might be worthwhile to confirm and improve this part of the model by finding data on industry migration under different conditions.

The aggregate mode shares do not change too much (percent by private auto is 83.7 in the trend scenario and 85.0 in the Beltway scenario, with most of that gain at the expense of non-motorized modes) but the total distance travelled increases significantly from 8.47 million km to 9.66 million km. The change in VKT is less dramatic, as the longer work trips are more likely to be made by HOV.

6.2.3 *LRT construction (Rail Scenario)*

6.2.3.1 Description of the Scenario

The Rail scenario includes the roadway projects of the Trend scenario, plus an extensive LRT construction program in between 2000 and 2005. There are 105 km of new track to complement the existing 34 km. The operating cost of private vehicles is increased by 30%, and a parking tax is introduced in the CBD representing an average surcharge of $4 for work trips and $1 for other trips.

6.2.3.2 Discussion of Results

In the Rail scenario the CBD is seen to lose employment as businesses relocate to other nearby zones to avoid the parking surcharge. The CBD gains residents, since commercial activities are no longer as willing to outbid residential activities.

The increased mobility over short distances in central zones (as compared to the trend scenario) allows a greater separation between population and employment just as in the Belt scenario. However the effect is much smaller with the rail scenario and only occurs to any degree in the most central zones where rail service is very good. Over longer distances and where rail service does not exist the increased private vehicle operating cost leads to shorter trips and more carpooling. As a result, total kilometers traveled drops somewhat and VKT drops significantly.

The mode split to transit is higher than in the base case scenario (3.14% vs. 2.33%). The extensive LRT construction allows transit to retain the same mode share as in 1990.

6.2.4 *LRT Construction in 2015 (Late Rail Scenario)*

6.2.4.1 Description of the Scenario

The Late Rail scenario has the same transportation infrastructure in 2015 as the rail scenario. The construction of the rail system and the imposition of charges on private vehicle use is delayed by 10 years, and so 10 years of land use adjustments do not take place in this scenario. A comparison of the Rail scenario to the Late Rail scenario shows how the land use component of the model influences overall results.

6.2.4.2 Discussion of Results

The land use patterns in the Late Rail scenario in 2015 are identical to the trend scenario, since land use patterns take time to adjust in MEPLAN.

Table 2 compares transportation characteristics between the Trend scenario, the Laterail scenario and the Rail scenario.

		Trend	Rail Construction	
			in 2015	in 2005
Mode Split (%) in 2015	private	83.73	83.39	82.27
	transit	2.33	2.66	3.14
	non motorized	13.95	13.94	14.59
Daily Person km of Travel (PKT) in 2015		8469000	8485000	8147000
Daily Vehicle km Travelled (VKT) in 2015		5925000	5860000	5297000

Table 2: Comparison between rail construction in 2015 and rail construction in 2005

Most of the shift in mode split from the Trend scenario to the Rail scenario only occurs when the patterns of interaction are allowed to adjust over 10 years. Almost all of the change in VKT is caused by changes in economic patterns.

6.2.5 *Transit Oriented Development (TOD Scenario)*

6.2.5.1 Description of the Scenario

The TOD scenario is a dramatic scenario involving land use policy changes and substantial investment in transit. In the year 2000 subsidies of 14% of year 2000 expenditures on rent are introduced in those zones where LRT will be constructed. These are offset by 30% surcharges in other zones so that region wide the effect is revenue neutral. In 2005 the LRT from the Rail scenario is constructed, but transit frequencies are doubled from the Rail scenario. In addition the value of waiting time is reduced by a factor of three to represent the effects of some form of sophisticated transit information system. In those zones where LRT is constructed the access time to LRT is reduced by three minutes to represent concentration of activity near LRT stations and a local para-transit service for accessing and egressing LRT.

6.2.5.2 Discussion of Results

The land subsidies attract substantial amounts of activity. Figure 8 shows where the growth occurs over the 25 years of the model. Almost all of the employment growth is attracted to zones with land subsidies, and many zones that do not have LRT service actually lose employment. Households are also attracted to the subsidized zones, but to a lesser degree. Figure 7 and figure 8 can be compared to see the differences between the Trend scenario and this TOD scenario.

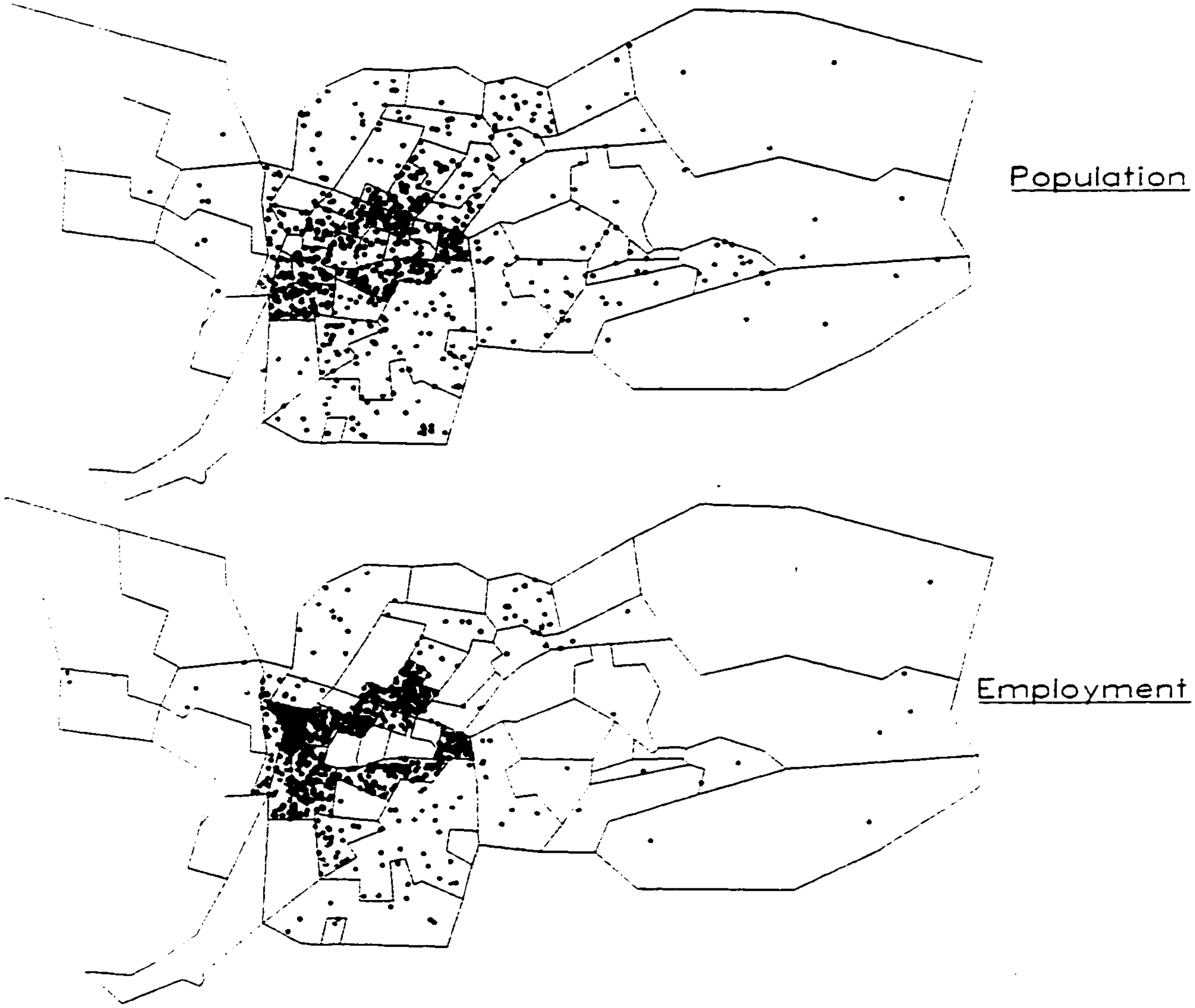


Figure 8: Maps of the allocation of increases in population and employment in the TOD scenario. A dot represents 500 households in the top map and 500 employees in the bottom map

The rents in the subsidized zones go up, and the rents in the taxed zones go down. This is because activities bid against each other to locate on the subsidized land. Hence most of the subsidies and taxes ultimately flow to the landowners.

It would be possible to bring the development to the LRT zones through the use of regulation rather than subsidies and taxes. Development permits might only be issued in the LRT zones. Such regulations will still disrupt the economy, however, and the model shows that the disruptions are approximately equivalent to a 14% subsidy in LRT zones and 30% tax in other zones. A possible future modelling exercise would be to take the land development pattern that is achieved in this scenario and apply it to another scenario using zoning regulations instead of subsidies and taxes. This would allow some direct comparisons between a tax and subsidy policy and an 'equivalent' zoning restriction policy.

The original policy plan for this scenario was to have all growth, including residential, occurring in the LRT zones in Transit Oriented Development. The simulation shows that the levels of subsidy and tax necessary to entice all new residential activity into those zones would be very high.

This simulation demonstrates the interactions and dependencies between different factors. The land subsidy in the CBD draws households into the CBD, and employment in the retail and education sectors also locates in the CBD to be close to the households. Other sectors are pushed out of the CBD. Each activity has a different consumption function for space and different dependencies on other activities, so the overall changes in activity location patterns can be quite complex.

This can be seen by looking at the kilometers traveled results. This TOD scenario gives a 15% lower VKT than the trend scenario, but the VKT for retail and for home-other trips actually increases.

The mode shifts are fairly dramatic - transit increases its mode share from 2.33% in the trend scenario to 9.36%. Most of this increase is from the SOV auto mode.

6.2.6 *Transit Oriented Development with Road and Parking Chargers (Pricing TOD)*

6.2.6.1 Description of the Scenario

The Pricing TOD scenario is identical to the TOD scenario, but with the addition of a 30% increase in automobile operating costs and a parking surcharges in the CBD of $5 for work trips and $1 for other trips. For work trips to some other zones near the CBD a parking cash-out policy was simulated as equivalent to a $1 parking surcharge.

6.2.6.2 Discussion of Results

The increased operating costs of private vehicles is a centralizing force, and in general both employment and households move towards the more central zones as compared to the TOD scenario. The parking surcharge in the CBD causes a distortion in this pattern, driving many businesses out of the CBD, and residential uses move into the CBD. Much of the employment moves into the Rancho Cordova suburban area east of Sacramento city towards Folsom.

The VKT drops substantially more. Table 3 shows a comparison of person kilometres travelled and VKT by scenario.

	Scenario					
Distance Measured (km)	1990	Trend	Belt	Rail	TOD	Pricing TOD
Person Travel by Network State						
Walk	63000	106000	97000	113000	103000	109000
Drive	2620000	4460000	3830000	3690000	3740000	3300000
LRT	29000	55000	55000	100000	209000	216000
HOV Auto	2090000	3670000	5490000	4010000	3590000	3560000
Cycle	45000	100000	82000	105000	101000	106000
Walk to transit	9200	11600	11400	16100	49600	57500
Bus	74000	69000	79000	103000	441000	594000
Total Person km (PKT)	4930000	8470000	9660000	8150000	8240000	7940000
Vehicle km (VKT)	3450000	5920000	6040000	5300000	5180000	4720000

Table 3: Kilometres of travel for different scenarios

The Pricing TOD scenario shows the lowest amount of travel and the lowest VKT of all the scenarios, but the VKT is still substantially higher than the value for 1990.

The Pricing TOD scenario also has the lowest mode split to private automobile (75.6%) with 10.9% of trips by transit and 13.5% by non-motorized modes.

In spite of the emphasis on rail, table 3 shows that it is the bus service that carries the most additional passenger kilometres. The doubling of frequencies and the reduction in waiting time disutility make the bus service attractive and viable for many trips.

6.2.7 *81ZN - Unconstrained zone 81*

In an initial run of the model extensive development was predicted in a far north-east zone (zone 81) with substantial amounts of vacant land. The development occurred to such an extent that a central place developed that was essentially self sustaining economically. This demonstrates that MEPLAN's representation of the positive feedback of agglomeration economies is sufficient to model "extreme sensitivity to initial conditions". Thus it seems likely that MEPLAN could represent the chaotic nature of urban systems as described by Allen (1997).

This node developed in the model because the model was not set up with significant detail in the zone to adequately represent the congestion and growth in intra-zonal travel time that would constrain such development. When this was corrected (in a somewhat artificial manner) the extreme growth no longer occured.

The 81ZN scenario is this initial run *without* the correction to zone 81. In future research this scenario could be used to investigate MEPLAN's ability to represent the chaotic dynamics of the urban system.

6.3 Implications of Results

The simulation results for the various scenarios show the capabilities of the model, and provide specific insight into the likely effect of policies in the Sacramento region.

6.3.1 Modelling Issues

6.3.1.1 Strengths and Weaknesses of the Modelling Framework

The market based nature of the modelling framework allows it to represent many well known and important characteristics from urban economics, including:

- the bidding process for land and the displacement of some activities by others willing to bid more, such as the displacement of residential from the CBD area in the Trend scenario,
- space consumption elasticities, where higher prices encourage higher densities,
- higher prices where scarcity exists, such as in Yolo county where most land is zoned only for agriculture.

The land market model also allows explicit representation of the effects of tax and subsidy policies, as in the TOD and Pricing TOD scenarios. This is a very important feature since it allows the model to be used for analysis in terms of the actual policy element being considered.

The allocation of development to zones is endogenous to the model. That is, price and other variables from one time period guide the location decisions of developers. This is an important part of forecasting, as developers are the actors that build most of the physical infrastructure of cities.

The quasi-dynamic nature of the framework allows for investigations into processes of change and the longer term effects of policy. The people, businesses and institutions in an area will take time to respond to policy changes, and a better understanding of when these reactions will occur is important so that planners are not unduly distracted by immediate short term results. For instance, the comparison between the Rail and the Late Rail scenarios shows that when a transit-focused transportation system is first instituted it is unlikely to have a large effect on VKT, but over time as activity patterns adjust to take advantage of the new possibilities VKT can drop considerably. The longest term effects are in development patterns, and the model's behaviour over time periods longer than 25 years might show larger effects.

The framework uses "factors" to represent the different aspects of economic activity, and each factor is homogeneous except for the probabilistic nature of the logit models that are used to simulate choices. The framework is thus an aggregate model that uses the theory and parameters from disaggregate models. This limits the ability to model certain unique situations. For instance in the Sacramento model the post secondary students at UC Davis live in all different types of households but exhibit behaviour that is different from other people in those households. To properly represent this would require segregating household types by the number of students in them, leading to an excessive number of household types. Models that work with samples of actors, such as microsimulation models, can easily handle this. But such models require sample data to work with, and aggregate data is usually easier to obtain, especially for business and institutional factors.

6.3.1.2 Strength and Weaknesses of the Model Design for Sacramento

The specific design of this model for the Sacramento region was driven largely by the data that were available and by characteristics of the Sacramento region that became apparent during calibration.

The City of Davis and the University of California at Davis have unique characteristics. Trip generation rates are higher, bicycle use is very high and the economic connections between students, their parents and the State of California are complex and difficult to represent. The targets relating to the City of Davis were strongly influencing the parameter estimation process. Thus specific changes were made to represent the City of Davis accurately without compromising the model's accuracy in the rest of the region. The relationship between UC Davis and the students in Davis was specified separately (exogenously), rather than allowing the spatial economic model to simulate it. The land used by the students is still endogenous, and the representation of the land market in Davis can still represent the shortage of land caused by agricultural zoning and the resulting higher rents.

A substantial number of home-other trips were destined to the service sector of the economy, as were a substantial number of inter-industry trips. To closely match the trip length data, the portion of this sector that serves households needed a much higher production allocation parameter than the portion that serves other industry. The services sector was split into two factors, OFSRV-RES and OFSRV-IND, to allow the household-serving OFSRV-RES factor to locate closer to the population it served.

6.3.2 *Policy Issues*

The Rail, TOD and Pricing TOD scenarios all show that various combinations of measures designed to reduce private vehicle use and encourage transit use can be effective. Reducing VKT even further beyond what is accomplished in the Pricing TOD scenario would probably require more substantial pricing to discourage automobile use, since the improvements to transit and the land use measures are already quite substantial.

Even in the Pricing TOD scenario that is designed to make transit most effective, the non-motorized modes have a higher mode-share than transit. This suggests that policies and facilities to encourage walking and cycling might be most effective in reducing SOV use and VKT.

If major new directions in transportation policy are pursued it is important to allow activity patterns and, ultimately, land development time to adjust. The initial shift in behaviour is likely to be followed by a more comprehensive shift over the years.

In the Rail scenario, the parking surcharge for work trips in the CBD had a fairly substantial effect on most industrial sectors, driving them out of the zone and driving down land prices. This policy is usually suggested as a measure to encourage transit use where transit is most viable, but the simulation suggests that the effect on activity location is likely to be far more important than the effect on mode choice. The policy might be more acceptable if the revenue from the surcharge were reinvested in the zone to make it more attractive in other ways. A further scenario run could investigate this possibility.

The average trip length of work trips was forecasted to grow in future scenarios (except in the Pricing TOD scenario.) This represents the normal growth in home and workplace opportunities that occur when regions grow. This is not necessarily bad, in fact it could be argued that this is completely intertwined with the general benefits of a larger economy. Education trips did not become longer, suggesting perhaps that a larger economy does not benefit education.

7 Firm Location in the MEPLAN Model of Sacramento

7.1 Introduction

This chapter describes the use MEPLAN in the analysis of firm location choice and the movements of firms in the policies evaluated using the Sacramento model. It describes the mechanisms in the model that allow for realistic aggregate assignment of firms to zones, and the resulting impact that firm location has on the entire modelling system.

MEPLAN uses a form of logit function to allocate production activities to zones. Interdependencies in the Social Accounting Matrix, together with trip and land cost data and usage rates, generate a utility function for the attractiveness of purchasing a given sector's output from a given zone.

Results from the model show the various features and strengths of the model, while pointing out some weaknesses and potential pitfalls. The strengths include the realistic representation of current and continuing patterns, the close linkages between different industry and household types and the market based nature of the model. The weaknesses include the aggregate nature of the model and the possibility of difficulties in establishing realistic alternative specific constants.

7.2 Economic Activity Allocation in MEPLAN

The Social Accounting Matrix in MEPLAN contains the information on the relationships between different 'factors'. There are four factor types in a typical MEPLAN model: industries, households, floorspace (buildings) and land. Industries consume output from other industries in the production process, along with labour from households and floorspace. (Land is not usually consumed directly by firms, but instead floorspace 'consumes' land when it is constructed.) The outputs from industries are consumed exogenously as exports, and endogenously by other industries and by households. These input and output relationships directly influence firm location in MEPLAN.

The mathematics of the MEPLAN framework is described in chapter 3. This section repeats much of the same information, but with some simplifications, in a different order and with an emphasis on firms.

MEPLAN is a market based model, and so costs and prices are important. The cost of production is usually the cost of all the inputs to production. Thus for a firm, the cost of production in a given zone is the cost of supplies from other firms, the cost of labour and the cost of floorspace. Each such cost is the amount of consumption multiplied by the price, and so the total cost of producing a unit of *m* in a zone *j* is (see also equation 9):

$$^{T}b_j^m = \sum_n a_j^{mn} \cdot {}^{T}p_j^n \qquad (13)$$

where:

m and *n* = indices representing factors. In this case *m* represents the production of the firm under consideration and *n* represents the other industrial goods, industrial services, labour and floorspace that the firm needs to consume,

$^{T}b_j^m$ = the cost of producing a unit of *m* in a zone *j*,

a_j^{mn} = the technical coefficient describing the amount of *n* needed in the production of one unit of *m* in zone *j*,

$^{T}p_j^n$ = the market price of a unit of *n* in the zone *j*.

For most factors, the market price of a factor *n* in a zone *j* is the weighted average cost of producing the factor in each zone *i* plus the cost of transporting it to zone *j* plus any excess profits, with the weighting being by interzonal trade volumes (see also equation 7):

$$^{T}p_j^n = \frac{\sum_i t_{ij}^n \left({}^{T}b_i^n + P_i^n + \delta_{ij}^n \right)}{\sum_i t_{ij}^n} \qquad (14)$$

where:

t_{ij}^n = the amount of trade in factor *n* from zone *i* to zone *j*,

P_i^n = the amount of excess profit made when producing a unit of factor *n* in zone *i*,

δ_{ij}^n = the money cost of transporting a unit of factor *n* from zone *i* to zone *j*.

Normally the excess profit term is kept at zero, implying that new production is attracted quickly by any hint of excess profit. If excess profits are non-zero it represents a shortage or surplus, and this can be implemented in MEPLAN by constraining the amount of production in zones. If consumption is elastic with respect to price the demand will decrease as profits increase; and MEPLAN will calculate the excess profit that is necessary to ensure the market clears. Space factors are usually represented this way: since space is non-transportable ($t_{ij}^n = 0$ for all $i \neq j$) and can only be created over longer periods of time by developers the markets for space in any one zone are not modelled as being in this type of equilibrium.

Equations 13 and 14 show that production costs are functions of market prices and that market prices are functions of production costs. If profits are known, the technical coefficients fixed and the travel costs fixed, then the only unknowns are the patterns of trades t_{ij}^n. These trade volumes are based on the disutility of acquiring n from zone i and consuming it in zone j, $-u_{ij}^n$, given by (see also equation 6):

$$-u_{ij}^n = {}^T b_i^n + P_i^n + {}^Q \varepsilon_i^n + d_{ij}^n \tag{15}$$

where:

${}^Q \varepsilon_i^n$ = a zone-specific disutility associated with producing n in i,

d_{ij}^n = the disutility of transporting n from i to j.

This utility formula is deterministic. Used directly, it would produce 'sharp boundaries', where categories of implicitly identical firms occupied areas with distinct limits. In reality, individual firms are unique, and a probabilistic model is needed to simulate variation within industrial factors without simulating each individual firm. To make the model probabilistic a random variable is added to the disutility. The random variable accounts for all other effects not included in equation 15 and also accounts for random variations in the costs, profits, zone specific disutilities and travel disutilities for individual sites and firms. Assuming that the random error occurs with a Weibull distribution, and that the error is for each site within each zone, leads to the logit formula for trade volumes (see also equation 6):

$$t_{ij}^n = {}^T c_j^n \frac{\exp\left[\lambda^n \cdot \left(u_{ij}^n + s_i^n\right)\right]}{\sum_i \exp\left[\lambda^n \cdot \left(u_{ij}^n + s_i^n\right)\right]} \tag{16}$$

where

${}^T c_j^n$ = the total amount of n consumed in j.

λ^n = dispersion parameter that is inversely related to the standard deviation of the error variable,

s_i^n = a 'size term' proportional to the log of the number of sites available to n in zone i.

The new unknowns introduced are the total consumption of each factor n in each zone j, and these can be found since the amount of consumption is the amount that is needed by all the production in the zone (see also equations 1 and 2)

$$^{T}c_j^n = \sum_m \left({}^{T}g_j^m \cdot a_j^{mn} \right) \tag{17}$$

where

$^{T}g_j^m$ = the total amount of m produced in j.

The production in each zone is the sum of the trades from that zone, plus any exogenous production corresponding to Lowry basic activity

$$^{T}g_i^n = \sum_j t_{ij}^n + {}^{Q}g_i^n \tag{18}$$

where:

$^{Q}g_i^n$ = the exogenous production of n in zone i.

The above set of equations constitutes the major part of the equilibrium land use model, and the equations are solved simultaneously by iteration. Other equations not shown here but described in chapter 3 include equations for allowing the technical coefficients to vary with price, and the equations for updating the profits or zone specific disutilities to match constraints. There are other forms available for some of the equations, especially for allowing household consumption patterns to be consistent with household utility maximization, and some equations have additional terms not shown here. For complete details refer to the MEPLAN User Reference Manual (ME&P, 1994).

The attributes of travel enter the location decision twice. In equation 15 the disutility of transporting the firm's production to the customer appears, and this affects equation 16 so that customers are more likely to purchase from close zones (or, equivalently, firms are more likely to supply to close zones.) But the production cost term in equation 15, $^{T}b_i^n$, comes from equations 13 and 14, and so includes the cost of purchasing and transporting the various inputs to production. That is, firms are also likely to locate in zones that are cheaper to them, and their costs include all inputs and (for transportable inputs) the costs of transportation. Thus firms will want to locate to be closer to their suppliers and their labor.

In this way MEPLAN can represent the classic trade-offs in location choice:

- The trade-off between being close to suppliers and close to customers, and
- The trade-off between transport costs and land costs.

7.3 The design of the Model for firm location in Sacramento

Figure 4 shows the design of the model for Sacramento, and chapter 5 describes the design and the design process. This section describes the design from the perspective of firm location.

7.3.1 Industrial Factors

The division of the economy of Sacramento into factors was done according to the categories of employees and industry that were provided. There are seven categories of employees, and service employees were eventually divided into two types during calibration. This gave the first eight factors in the MEPLAN model:

- AGMIN — agriculture and mining
- MANUF — manufacturing
- OFSRV-RES — services and office employment consumed by households
- OFSRV-IND — services and office employment consumed by other industry
- RETAIL
- HEALTH
- EDUCATION
- GOVT — government

These were seen to be appropriate factors because data existed on the numbers of each of these types of employees in each zone. By making the model correspond to these employees it was possible to explicitly represent the spatial allocation of different types of labour in different areas of the Sacramento region. These were related to the Social Accounting Matrix to determine the inter-industry connections. Education employees were not related to the Social Accounting Matrix (as described in Chapter 5).

This accounted for most of the SAM. The remaining sections that had not been related to employment data were:

- PRIV EDU private education
- TRANSPORT commercial transportation
- WHOLESALE

These sections of the economy are represented directly by three additional factors in monetary units.

Thus there are seven factors in the model corresponding to both the employment data and the SAM: AGMIN, MANUF, OFSRV-RES, OFSRV-IND, RETAIL, HEALTH and GOVT. Three factors correspond only to the SAM: PRIV EDU, TRANSPORT and WHOLESALE. One category, EDUCATION, corresponds only to employment data.

The consumption of labour by industry was not included in the SAM, so conversions were needed for the seven factors that represent both employment data and SAM data. The total size of the sector and the total number of employees in the sector were used to determine the amount of output produced by one employee. This was made implicit in the model by using 'one employee's output' as the units for measuring these factors' production. The consumption of households by employees (that is the supply of labour by households) required data on the number of households, by household income, needed to supply one employee in each industry. The number of households 'in' each industry was available, and dividing these numbers by the total employees in each industry gave the coefficient that indicated the number of households needed to supply one employee.

Industries and households consume space at different rates and have different price elasticities, giving seven land use factors:

- AGMIN LU   land used for agriculture
- MANUF LU   land used for manufacturing
- OFSRV LU   land used for services and office employment
- RETAIL LU   land used for retail
- HEALTH LU   land used for health
- EDUCATION LU   land used for education
- GOVT LU   land used for government
- RES LU   land used by residences

These allow for taxes, subsidies and regulatory zoning that affect only certain industries in certain zones.

7.3.2 *Exogenous Demand*

The SAM listed the total exports from the region, but the geographic arrangement of the exporting production was not available. The 'exporting employees' were allocated to all zones in proportion to the total size of that factor in each zone. The exogenous households (the retired and unemployed) were also allocated to zones in proportion to the total number of households (by income) in each zone. Davis student households and the staff at the University of California were made exogenous to account for various unique characteristics of the university population of that city.

The long thin matrix just above the large matrix in figure 4 shows the exogenously demanded activity. The 'r's in the top matrix for the industry and household factors show that the growth in exogenous activity is allocated to zones in the incremental model according to costs and existing distributions. Specifically, new exogenous activity locates in zones inversely proportional to production costs, ${}^{T}b_i^n$, but proportional to that factor's arrangement in the previous time period.

7.3.3 *Interface and Transport Model*

There is no 'goods movement' flow in the Sacramento version of MEPLAN at present, although MEPLAN models usually include this and the Sacramento model has been designed to allow this to be implemented at a later date. This is an important limitation, as currently the relationships between industries are calibrated to the costs and times for passenger travel. Adding a flow for freight movement would improve the general accuracy of the model's firm location choice and would allow transportation conditions for freight to be investigated separately from transportation conditions for passengers.

In the transport model, five modes are available, and each mode can consist of several different types of activity ('states') on different types of links.

7.3.4 *Calibration*

The heuristic parameters that have the most effect on firm location choice are $^{Q}\varepsilon_i^n$ (the alternative specific constant for factor *n* in zone *i*), and λ^n (the dispersion parameter in the logit allocation model for factor *n*). These were estimated using the calibration process described in chapter 9.

7.4 Firm location in the Scenario results

7.4.1 *Minimal Construction (Trend Scenario)*

7.4.1.1 Description of the Scenario

The Trend scenario is the 'Current Plan Constrained' option from the 1996 Metropolitan Transportation Plan for Sacramento (DKS, 1994). It involves little transit expansion, a modest amount of HOV lane construction and a respectable amount of road building to 2005, with very little expansion beyond 2005.

7.4.1.2 Discussion of Results

Figure 9 shows how the increase in activity for each economic factor is allocated to zones, and figure 6 shows the zoning system. The increases are not uniform, but arrange themselves around the urban area according to the various costs and prices that are simulated. The patterns are complex and difficult to see without color maps, but some observations can be made.

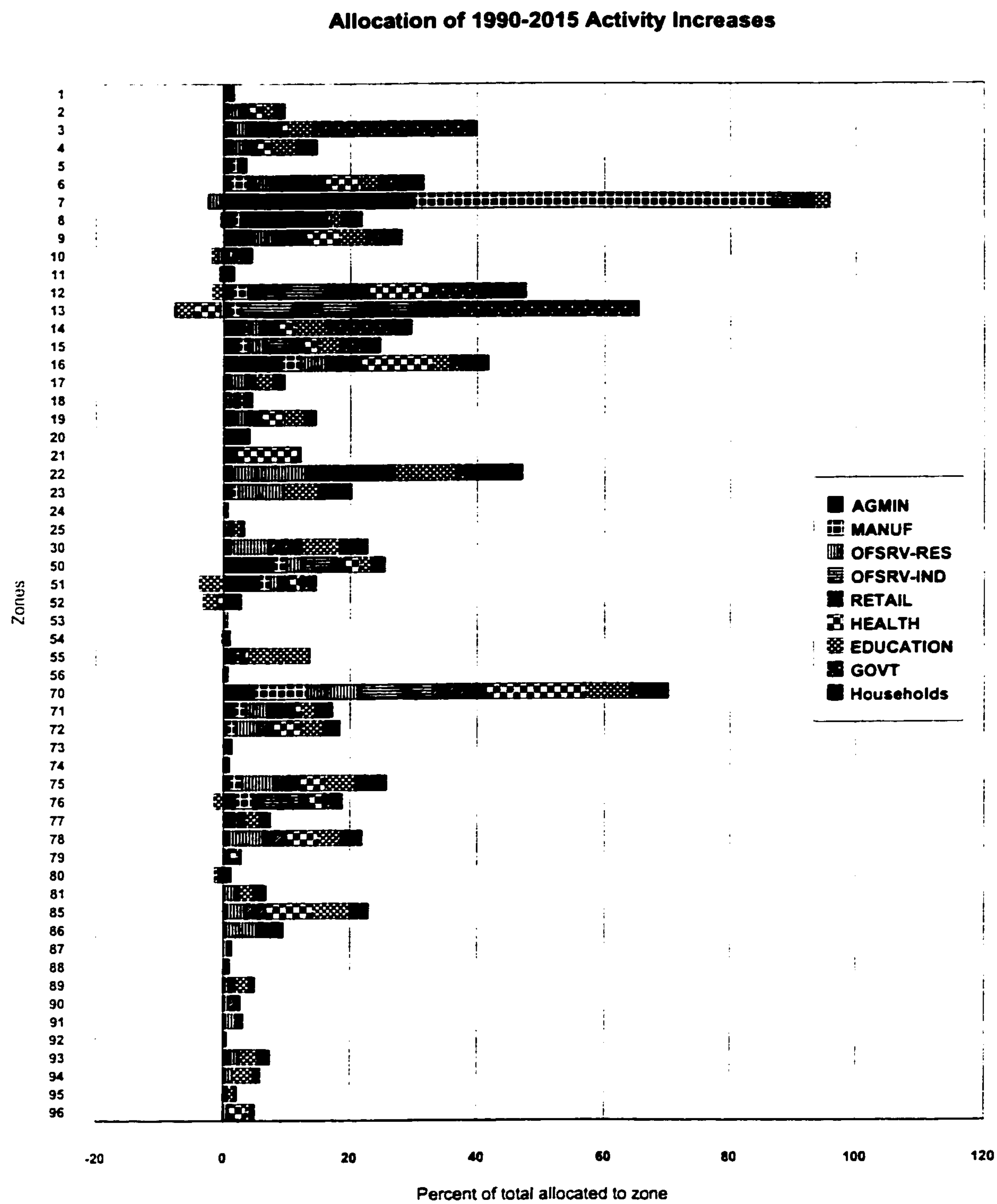


Figure 9: Sector breakdown of the allocation of increases in population and employment in the Trend scenario

The growth in government services is largely allocated to zone 13 (the CBD) and zone 3 (east of the CBD). The model allocates this employment to these zones because of the values of the alternative specific constants. Governments have their own reasons for locating in certain locations, many of them are historic or political and cannot be represented in the model's costs. These effects are captured in the alternative specific constants, which act as 'catch all' terms. The alternative specific constants are determined during calibration, when they are adjusted until the distributions in the base year are reproduced. This is what allows the model to make this realistic projection of government activity patterns.

Figure 10 shows, for each zone, how each employment sector's share of the total employment changed between 1990 and 2015. In zone 13 and zone 3 the composition of the employment by sector did not change too much (even though the total employment increased substantially).

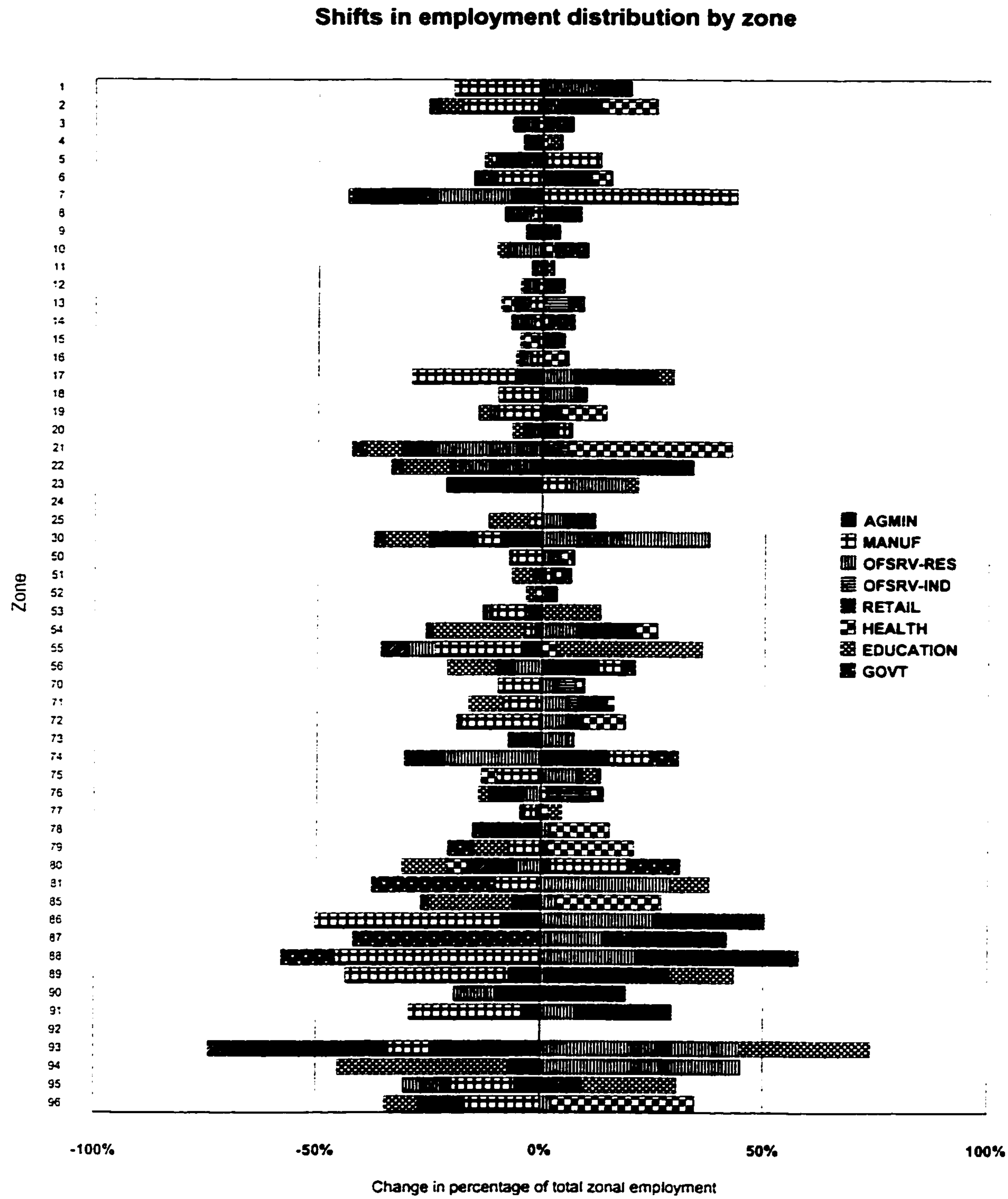


Figure 10: Shifts in the composition of employee types in zones, for the Trend scenario 1990-2015

The changes in zone 7 (immediately north of the CBD) are enlightening. Figure 9 shows that a substantial amount of the increase in manufacturing activity is allocated to zone 7. Figure 10 shows that this increase in manufacturing is at the expense of service employment. This is caused by the very high prices for manufacturing land in zone 7 in the calibration year. The amount of manufacturing activity that was willing to locate in zone 7 in 1990 in spite of these very high prices suggests a certain attractiveness and scarcity of manufacturing land in zone 7 and/or certain historic trends. In fact, this is the old Southern Pacific Rail yards, where a unique situation has arisen from the combination of changing technology, environmental constraints and speculative land holdings. Within the model, development is attracted to zone 7 by the high prices, and manufacturing activity follows development to locate in this attractive zone. This shows the importance of good price data - an abnormally high price in the calibration year can influence the model's predictions substantially.

Zone 22 (Southeast) shows substantial growth in many sectors (figure 9) because of its size and location; but retail uses begin to predominate over other sectors (figure 10). The need for retail comes from the growth in population in this zone and in surrounding zones. Population also requires education services, health services and office-services. These occur in the neighboring zones of 12, 17, 19, 21, 23 and 85. The distribution of health activity in the calibration year causes another deviation in the alternative specific constants. Zones 17, 22 and 23 have no health employment in 1990, and the calibration process assumes that no health activity locates in these zones because they are unattractive for this factor. This causes the future health employment demanded by the increase in population in these zones to be allocated to neighboring zones 12, 19 and 21.

This distribution of health activity is probably unrealistic. It is caused by a breakdown in the interpretation of the logit model for aggregate behaviour when values are small. The law of large numbers says that the share of decision makers choosing an alternative is equal to the probability of any one decision maker choosing the alternative if a large number of cases are considered. MEPLAN uses this property throughout the model, even when the law of large numbers is not accurate. In this case, since the share is zero, the alternative specific constant is assigned a value to make the probability zero.

The problem is not insurmountable. The calibration mechanism could avoid excessive alternative specific constants, or at least flag them as potentially dangerous. A better method could be used to determine the alternative specific constants: local knowledge of the characteristics of zones could identify zones that should have similar alternative specific constants, and zones with enough activity to make the law of large numbers accurate in the calibration year could be used to estimate alternative specific constants for all similar zones.

A larger discussion of the nature of zone specific constants and how they can be estimated is contained in section 9.3.1.

Figure 9 shows that in some zones (for example 7, 10, 13, 51 and 52) certain activities actually decrease in amount during the 25 years of the model. In the CBD (zone 13) there is a slight decrease in the amount of health employment, education employment and households. In the same zone there is a large concentration of additional employment for services to industry (OFSRV-IND) and of government employment. This represents a classic property from urban economics where some activities are able to outbid other activities, raising land prices and driving out these other activities.

7.4.2 *Extensive Highway Construction (Beltway Scenario)*

7.4.2.1 Description of the Scenario

The Beltway scenario involves substantial construction of new roadways and HOV lanes. This new construction is introduced in 2005, and so affects land use in 2010 and 2015.

7.4.2.2 Discussion of Results

Figure 11 shows the shifts that occur under this scenario relative to the trend scenario. An important change is the increasing importance of zone 12, the Rancho Cordova commercial node east of the City of Sacramento and west of Folsom. The losers of commercial activities are more distant zones, including the cities of Auburn (zone 76) and Folsom (zone 6). These peripheral cities are a little more like 'bedroom communities' in this scenario, as the roadway expansion allows industry to locate further away from the households that it serves and employs. For instance, retail activity can shift from local commercial to 'big-box' retailing as a result of increased roadway capacity.

It is important to note that the shifts in commercial activity caused by this infrastructure is quite substantial, and orders of magnitude greater than the shifts in residential patterns that occur.

Figure 11: Shifts in activity by zone and activity type for the Belt scenario, as compared to the Trend scenario

7.4.3 *LRT Construction (Rail Scenario)*

7.4.3.1 Description of the Scenario

The Rail scenario includes the roadway projects of the Trend scenario, plus an extensive LRT construction program in between 2000 and 2005. There are 105 km of new track to complement the existing 34 km. The operating cost of private vehicles is increased by 30%, and a parking tax is introduced in the CBD representing an average surcharge of $4 for work trips and $1 for other trips.

7.4.3.2 Discussion of Results

In the Rail scenario the CBD is seen to lose employment as businesses relocate to other nearby zones to avoid the parking surcharge. The CBD gains residents, since commercial activities are no longer as willing to outbid residential activities.

7.4.4 *Transit Oriented Development (TOD Scenario)*

7.4.4.1 Description of the Scenario

The TOD scenario is a dramatic scenario involving land use policy changes and substantial investment in transit. In the year 2000 subsidies of 14% of year 2000 expenditures on rent are introduced in those zones where LRT will be constructed. These are offset by 30% surcharges in other zones so that region wide the effect is revenue neutral. By 2005 the LRT from the Rail scenario is constructed, and transit frequencies are doubled. In addition the value of waiting time is reduced by a factor of three to represent the effects of some form of sophisticated transit information system. In those zones where LRT is constructed the access time to LRT is reduced by three minutes to represent concentration of activity near LRT stations and a local para-transit service for accessing and egressing LRT.

7.4.4.2 Discussion of Results

The land subsidies attract substantial amounts of development and activity. Activities bidding against each other raise the land rents, and the higher rents attract developers. The patterns of movements are complex, and relate to interdependencies between sectors of the economy. Education activity is affected directly by the subsidies, but also follows the changes in activity patterns of households since the households supply the students. The service sector for residents (OFFSRV-RES) and the retail factor seem to be most sensitive to the land subsidies.

7.5 Conclusions

The MEPLAN modelling framework provides a mechanism for simulating industry location dynamics. Consumption disutility functions for each factor and zone pair are used, with the disutility including the costs of production (which include related transportation costs and land costs), any excess profits, zone specific disutilities and the disutility of transporting the production. As such the model is able to represent the bidding process for land and the trade-offs between land costs, supply costs and transportation.

Industry is represented with broad factors in MEPLAN, and land as represented as zones. This simplifies the modelling considerably, but the aggregate treatment can not represent the unique nature of firms and the unique nature of location decisions.

The zone specific constants are important. These allow existing patterns that occur for reasons not endogenous to the model to be carried through to the future. But they are also 'dangerous' in that they can take errors that are unimportant in the calibration year and carry them through to future years, where they may be considerably magnified. Different ways of estimating and investigating these zone specific constants could be pursued; this is discussed in more detail in Chapter 9.

The movements of industry in the scenarios for Sacramento are quite substantial, and are much larger than the movements of households. This demonstrates the critical importance of considering firm location choice decisions when planning. It would seem to be at odds with the greater emphasis on residential location choice behaviour in the literature (Hunt *et al*, 1994; Hunt, 1997), which has perhaps arisen because of the greater availability of data concerning household movements and the comparative homogeneity of household categories. The spatial activity response of models could be severely biased if only the residential component is considered.

The magnitude of the movement of industry suggests that it is important to properly calibrate the underlying relationships that cause this movement. In the MEPLAN model it is important to ensure that the willingness to travel vs. the willingness to change locations is properly represented. At the aggregate level, the trip length distributions for inter-industry trips are the primary indication of this trade-off. Thus, when the overall modelling system is constructed and run the parameters should be adjusted so that the trip lengths match the observed data. The costs of travel and of location and firm's willingness to trade-off between these costs should be further investigated, perhaps using disaggregate data and "extra" models that are more sophisticated than the ones that can be reasonably included in the overall urban model.

In various scenario runs the interdependencies between different industry types proved very influential - activity locates so as to be favorably positioned with regard to inputs and consumers, and the differences in relative consumption by different industries and households leads to closely linked activities 'following each other' around the urban area. Cross movement of different types of industry would be partially constrained in the real world because of the need for different types of building stock and the long time frame needed to redevelop that stock. The Sacramento model includes redevelopment in the equilibrium model. If and when further information on floorspace becomes available it might be appropriate to move the representation of redevelopment to the incremental models and to develop a wider range of floorspace categories.

In an ideal situation, the following data would available for calibrating an accurate firm location choice model within MEPLAN:

- the spatial arrangement of employment and activity,
- the spatial arrangement of exogenous activity,
- a standard Social Accounting Matrix,
- land and space consumption rates,
- the physical arrangement of various types of land and buildings,
- the prices of various types of land and buildings,
- the relationships between economic links and trip rates or freight volumes, and
- trip cost and disutility for different trip types and modes.

Not all of these data were available for the Sacramento model; this project has demonstrated that the framework can be used to develop a useful model of firm location dynamics in a United States situation using existing data sources.

Part 3.

Parameter Estimation

8 Parameter Estimation Strategies for Large Scale Urban Models

8.1 Chapter overview

Part 2 of this dissertation described the construction of a land-use transportation model of the Sacramento region and the use of the model for policy analysis. The parameters were estimated using the ad-hoc techniques typical for such modelling efforts, but with a manual calibration of the overall system.

These ad-hoc techniques are not normally justified through statistical theory — they are applied by modellers and generally accepted as "best practice" even though there is not a common understanding of why or whether these techniques are acceptable.

Part 3 of this dissertation, beginning with this chapter, directly addresses the hypothesis that a systematic overall calibration is necessary and desirable for large scale urban models. First, in this chapter, some different techniques for parameter estimation in large scale urban models are presented, together with a gradual development of the statistical theory of parameter estimation in such models. The advantages and disadvantages of the different techniques are presented.

Large scale urban models are often subdivided into simpler submodels. The parameters of these models can be estimated using approaches that differ in whether the full modelling system is run during an estimation procedure and whether that overall estimation is performed simultaneously with the estimation of the individual submodels. There are also ways in which extra data or extra models can be used to further inform parameter values. Five different techniques are presented ("Limited view", "Piece wise", "Simultaneous", "Sequential" and "Bayesian sequential"), and concurrently the statistical theory necessary to justify each technique is described. The practical advantages and disadvantages are discussed in this chapter, and each technique is illustrated using a simple nested logit model example. The ideas and examples should illustrate the importance of an overall calibration.

8.2 Alternative Parameter Estimation strategies

A modular modelling system consisting of a series of interconnected submodels is shown in Figure 12. Various inputs are shown coming from the edges of the figure and attaching to the submodels at the dots. Model outputs are calculated by submodels and are drawn as arrows going to the edge of the figure. There are also connections between submodels, representing the data flows within the modelling system. Some model outputs are drawn from the same point as an internal data flow, showing that the same data can serve as both an output and a data flow.

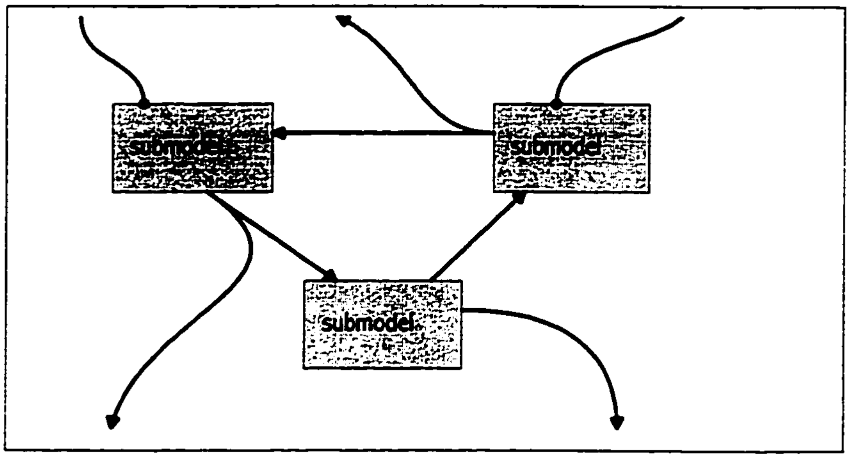


Figure 12: A Complex Modelling System

The submodels have a certain degree of independence from each other, and the degree to which they are treated as independent when estimating parameters leads to the different strategies.

8.2.1 Limited View Approach

With this approach the individual submodels are ignored during parameter estimation. The modelling system is run with the inputs set to observed values and the parameters are adjusted until the modelling system's outputs closely match corresponding observed values.

Ideally, a statistical approach is used, where a likelihood function (the probability that the model will predict the observed values) is maximized. The model is not expected to reproduce the actual data as outputs, for three possible reasons:

- The system itself is inherently random. This is certainly the case for the human behaviour that determines much of an urban system. This is probably the most important source of variation.
- The model is a simplification of reality that is not expected to be entirely accurate. Cities are complicated systems and models can not represent all of the complicated relationships that occur in reality.
- There are measurement errors in the data. This is true for all data, but in urban systems modelling the magnitude of the errors in the data are likely unimportant compared to the two other sources of variation listed above.

Since all of these occur in urban systems, the models should be probabilistic. The outputs of a probabilistic model are probability distributions. Yet in practice the probabilistic nature of models is often glossed-over by referring to the expected value of each probability distribution as a "predicted value" or "modelled value".

This can lead to misunderstandings and misuse of model outputs in practical work, so it is useful to reiterate. Probabilistic models produce, as outputs, probability distributions. The parameters are estimated by finding parameter values that cause these distributions to closely overlap the range of observations. Mathematically this is accomplished by maximizing the likelihood function. The likelihood function is the probability that a sample from the model's distributions will reproduce the observed values within some fixed tolerance (for continuous variables) or exactly (for discrete variables) (Bard, 1974). If the model is run in a way that produces a sample from the model's distributions then these outputs would constitute an *experiment* that would share many characteristics with the observed values that come from the real urban system, which can also be viewed as an *experiment* (Naylor, 1971). In practice, however, the expected values of the distributions are often reported as "model outputs", and these outputs are practically and theoretically different than what will actually occur.

The complexity and non-linearity of an interconnected modelling system leads to difficulty in formulating the likelihood function. This often leads to a least-squares approach, which implicitly assumes that all outputs have independent normal distributions. The non-linear least squares estimation problem is usually performed using a heuristic search algorithm (Bard, 1974).

8.2.1.1 Advantages

Estimating parameters using this "limited view" approach has a certain appeal. Perhaps most importantly it produces parameters based on the premise that the model should be able to reproduce the observations that correspond to the outputs of the modelling exercise.

Further, the model runs used when searching for the best parameter values can be essentially identical to the model runs that are performed when using the model to make predictions. This allows the model-builder to concentrate more on how the model will be used in application instead of how the parameters will be estimated, and the same software can be used for both tasks. Specifically, the reduced form of the model can be used, where the output variables are calculated as a function of input variables (Bard, 1974) .

$$y = f(w, x) \tag{19}$$

where:

y	is the vector of output variables
w	is the vector of exogenous (input) variables
x	is a vector of parameters
$f()$	is the calculation of the output variables; often a complex numerical procedure.

Focusing on the entire modelling system is likely to reveal problems in the structure of the model, since structural problems often show up as a difficulty in achieving a goodness of fit in one set of outputs without compromising the goodness of fit in another set of outputs. The focus on model outputs also helps to confirm the software implementation of the model, as bugs in solving the reduced form of the model will show up as errors in the model outputs.

8.2.1.1 Disadvantages

The reduced form (equation 19) needed to run the model has disadvantages during parameter estimation. For urban system modelling the error in the measurement of the data is often negligible in comparison to the errors that are inherent in the model's equations themselves; yet the likelihood functions that can easily be formulated from equation 19 are based on assumptions that the errors are in the output variables $\mathbf{y}$. In contrast, the structural form of the model,

$$g(\boldsymbol{x},\boldsymbol{y},\boldsymbol{w}) = 0 \tag{20}$$

where:

$g()$ — are the structural equations of the model corresponding to the theories on which the model is based,

would have simple error distributions if the errors are caused by the simplifications made in the model. Thus, a reasonable, theoretically defensible set of probability distributions can be hypothesized for the error in these structural equations. In simple models, the relationship between $g()$ and $f()$ could be used to transform the error distributions caused by the mis-specification of equation 20 into distributions of the output variables in equation 19; but urban systems models are rarely simple models. The relative complexity of the transformation between structural form and reduced form makes it difficult to formulate probability distributions for the output variables consistent with structural errors.

The calculation of the output variables (i.e. the solution of the reduced form, equation 19) usually requires complex numerical procedures. The calculation of the errors in the structural form (equation 20) is, by contrast, a relatively simple matter. Thus estimating a model by minimizing the errors in the structural equations could be computationally much simpler (as well as theoretically more appealing) than estimating a model by minimizing the errors in the outputs.

The "limited view" approach does not include any extra data in the estimation process. These extra data are functions of some of the model's parameters, but do not directly correspond to the models' predictions. An important example is the lower level choices in nested logit model parameter estimation. In a nested logit model the choice from among the full set of alternatives is predicted when running the model. But in estimation, various conditional choices are also used to estimate parameters.

The top set of lines in figure 13 shows this using as an example the choice of destination and mode. Household characteristics are inputs into the location choice submodel, as are measures of travel disutility. The predicted location choices are compared with the observed location choices during estimation. These predicted location choices are inputs to the mode choice submodel, and the full choice of mode and location is compared with the observed choice of mode and location. During model operation the same process is used — household characteristics are model inputs, and mode and location choice are predicted. The "limited view" approach for calibrating nested logit models would only use this top set of lines.

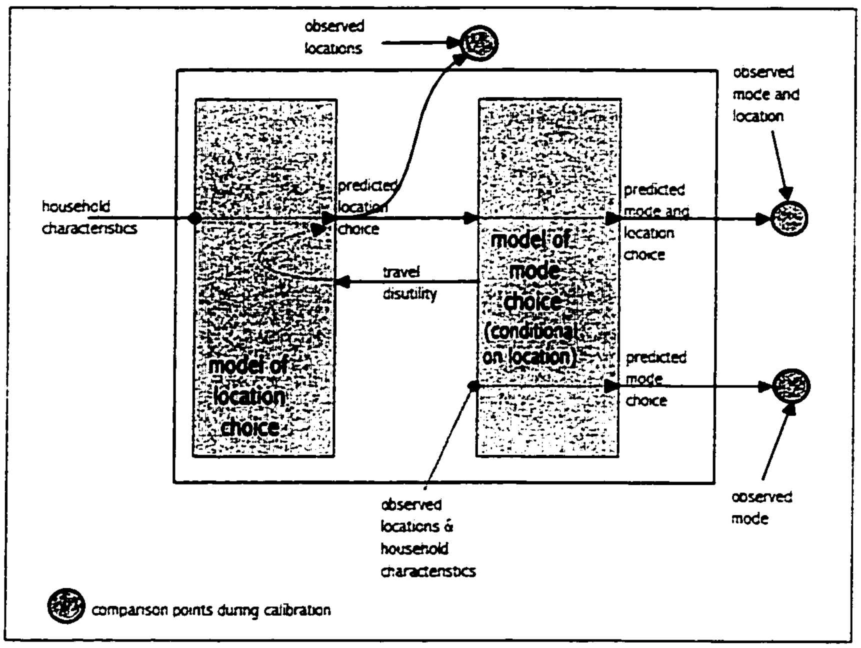


Figure 13: Example of using extra data during parameter estimation. The bottom data flow (grey lines) is "extra" since during model operation there are no observed locations

The bottom set of lines show another data flow that normally occurs during logit model estimation. The observed location choice and household characteristics feed directly into the conditional mode choice submodel (as shown by the grey line with a dot on the end), and the modelled choice of mode is compared with the observed choice of mode. This second data flow cannot occur during model operation because the locations are *predicted* by the model and not *observed*. These types of secondary data flows are crucial to producing estimates of parameter values that are close to the true values. The "limited view" approach cannot use these extra data.

The nested logit model example is important since logit formulations are used extensively throughout urban models and because the theory of how the submodels connect is well developed. But regardless of whether the model is logit or not, there are typically observed data that can further inform parameter values of an individual submodel, but that are extra to the modelling system.

This is especially the case for disaggregate data. The circular interactions between the various submodels usually means that all of the disaggregate data within the modelling system is synthesized by other submodels in the same way that the "predicted location choice" and "travel disutility" are synthesized in figure 13. Observed disaggregate data cannot be used unless these data are used in place of the synthesized data, which requires considering the individual submodels separately.

8.2.2 *Piece wise estimation*

In piece wise estimation, the connections between submodels are ignored. The parameters of each submodel are estimated based on the data that directly affect that submodel. This is shown in figure 14. After the parameters of each submodel are estimated the overall modelling system is constructed and run.

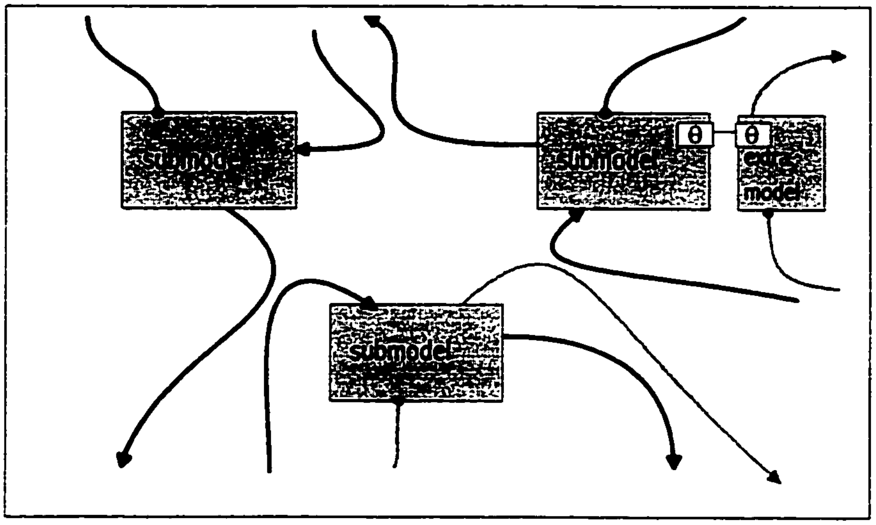


Figure 14: Piece wise estimation. θ represents shared parameters

8.2.2.1 Advantages

The primary advantages of considering each submodel separately are that it breaks the problem into more manageable pieces, and it allows more observed data to be considered since the synthesized data from other submodels can be replaced with observed data (if it exists).

Another important advantage is that it is easier to use extra data during the consideration of each submodel, as shown by the grey lines in figure 14. This is a very important advantage. In urban systems modelling entirely different data could be used to further inform parameter values. These could include targeted sample data, stated preference survey data, and even data from a different city.

Sometimes these extra data are in such a form that they cannot be used directly by the submodel within the modelling system. This is often the case when a submodel represents behaviour in aggregate for simplicity (e.g. zone-to-zone mode-split in a four step model) but there are disaggregate data describing the same relationships (e.g. individual mode choices from sample data). In this case an "extra model" is created to represent this behaviour, with parameters shared between the submodel and the extra model (figure 14). The parameter values can be taken directly from the extra model, or the two simple models can be estimated together simultaneously.

Models with parameters estimated in this way generally have more accurate submodels, since the submodels themselves are the focus during estimation. This can be an advantage if the submodels are to be used separately from the overall modelling system.

The individual submodels often correspond to well established theories and are often operationalized using fairly simple equations. These allow theoretically pure likelihood functions to be developed for individual submodels. The observed data used in estimation often have independent error distributions, further simplifying the likelihood function. All this often allows for fairly well developed and well accepted parameter estimation procedures. Hence the modeler can have more confidence in the statistical results (including goodness-of-fit measures and confidence limits on parameter values.) The well established theories that underlie these simpler models also give confidence in the reliability of the individual submodels.

8.2.2.2 Disadvantages

If observed data is not available to replace the synthesized data that normally connects submodels, then it is impossible to consider a submodel on its own. Thus it is not always possible to consider each submodel separately especially when data are scarce (e.g. in third world countries).

In fact, in typical nested logit models it is rarely possible to perform piece wise estimation because there is rarely any effort to directly measure the "expected utility" of choosing from a group of alternatives. In the example of figure 13, it is rare to have a measurement of the "travel disutility". It would be an interesting exercise to ask survey respondents to indicate their "composite utility of travel by all modes" and then use this data to consider the location choice submodel independently of the mode choice submodel. Such data could also be used to further inform the parameters of the mode choice submodel independently of the location choice submodel. Yet in practice this is not done for logit modelling.

The statistical properties of the input data, combined with the strong theory and simple equations that are often in each submodel, are an advantage when the models are separate. But when the submodels are combined into the overall modelling framework the input data for one submodel comes from the output data for another submodel. These synthesized data are subject to the systematic errors in the modelling system, so do not have the independent distributions that are usually assumed when considering the submodels separately. If dependent submodels are non-linear this could lead to a bias in outputs; causing a degree of inaccuracy in the entire modelling system (Cuthbertson *et al*, 1992).

In other words, if accurate submodels are combined into an overall modelling system *there is no guarantee that the overall modelling system will be accurate*. The way submodels are combined into an overall model will almost certainly violate the assumptions adopted when establishing the parameters of the individual submodels. The overall system needs to be investigated as a whole.

Finally, the whole purpose of comprehensive models is to see the "big picture"; to understand how changes in one aspect of the system can lead to changes in the overall system. These interactions are unlikely to be represented accurately without some effort to estimate parameters by considering the entire modelling system.

All this suggests that piece wise estimation is rarely an appropriate strategy for large scale urban models.

8.2.3 Simultaneous estimation

The most theoretically pure parameter estimation system for a complex modelling system is simultaneous estimation. Here the overall modelling system is run and its outputs are compared to various targets, but concurrently each of the individual submodels (and any "extra models" that can further inform parameter values) are also run to process the extra data available for those models. This is shown in figure 15, where all the data flows from both figure 12 and figure 14 are present. The goodness of fit measures from the individual models (which are likely to be likelihood measures from statistical theory) are combined together with the goodness of fit measures for the overall modelling system (which may also be likelihood measures, but are more likely to be quasi-likelihood because of the complexity of the overall system.) The overall likelihood (or quasi-likelihood) function is maximized to simultaneously determine all the parameters.

A well known example is the case of nested logit models, as shown in figure 13. Here the higher level location choice submodel is estimated with data on household characteristics, travel disutilities from the mode choice submodel (the log-sum term), and observed data on location choice. The mode choice submodel is estimated with observed data on mode choice. But the two submodels are estimated simultaneously by maximizing the joint likelihood.

8.2.3.1 Advantages

The simultaneous consideration of all the data by the entire modelling system can lead to more informed parameters. An important consideration is that any error estimates obtained for the parameters will be biased unless simultaneous estimation is performed (Amemiya, 1974).

Simultaneous estimation overcomes many data availability problems, since missing data can be synthesized from other submodels. In the nested logit model example of figure 13 the "travel disutilities" can be calculated by the mode choice submodel when its parameters are being estimated at the same time as the parameters in the location choice submodel.

Simultaneous estimation is especially appropriate when there is a theoretical reason why a parameter in one model should be identical to (or a function of) a parameter in another model (shown by the α and the θ in figure 15). With simultaneous estimation the parameter estimate can be influenced by all of the relevant data.

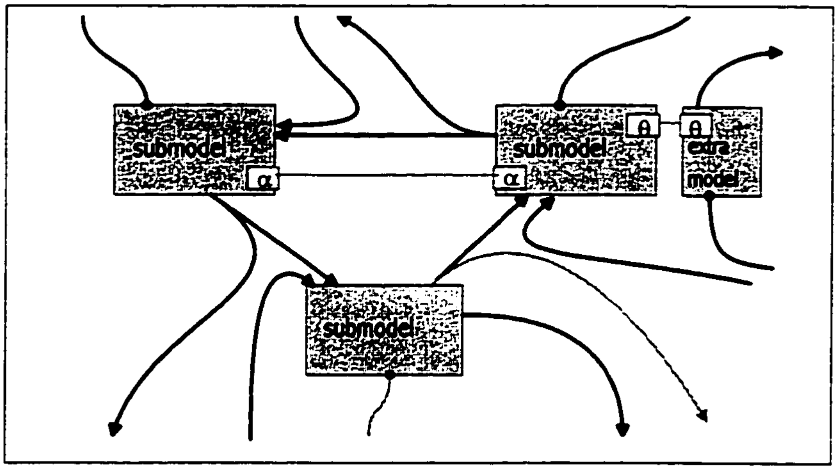


Figure 15: Simultaneous estimation. θ and α are parameters shared between different models

8.2.3.1 Disadvantages

The primary disadvantage of this method is its complexity. Each of the submodels must be run at every iteration, which can be computationally intensive. The primary reason for partitioning the modelling system is to reduce the complexity; to break the model into more manageable pieces. It would be difficult and computationally expensive to have all the models, (including "extra models") process all of their data (including "extra data") at every iteration of the parameter estimation process.

8.2.4 *Sequential estimation*

This technique combines piece wise estimation with the limited view approach to achieve many of the benefits of simultaneous estimation. The parameters of individual submodels are estimated, and then the overall model is considered. Various parameters are identified as being crucial to the higher level behaviour of the model, and these parameters are estimated by examining the entire modelling system.

In the nested logit example of figure 13, the parameters in the mode choice submodel are estimated based on the conditional choice of mode for a given location (the bottom data flow). The parameters of the location choice submodel are estimated by considering the primary data flow at the top of figure 13, using the travel disutility values from the mode choice submodel. Normally, the sequential estimation procedure for nested logit models is finished once the parameters of the highest level model have been estimated. But sometimes a final overall estimation is performed to ensure the aggregate predictions of the model correspond to various known aggregate quantities (for example, the estimation of various zone specific constants in Abraham and Hunt [1997])

Often the parameters are estimated using piece wise estimation, but the overall modelling system is "validated" once it is constructed. This refers to the process of running the overall modelling system and comparing various outputs to known values to see if the results are reasonable. If the results are not as good as expected parameters may be adjusted in an *ad-hoc* way to achieve a better fit, in which case a form of sequential estimation has occurred. Hence "validation" can be considered to be a way of avoiding the overall estimation or performing it in an *ad-hoc* way only if it is absolutely necessary. This is described by Popper (1959) who says that if, in a series of empirical tests of a model, no negative results are found but the number of positive instances increases, then our confidence in the model will grop step by step (Quoted in Naylor, 1971).

8.2.4.1 Selection of higher level parameters

In urban systems modelling, the situation is not usually as clear as in simple nested logit models. The theories of how individual submodels should influence one another are not as well developed. Hence the parameters to be estimated at the higher level are chosen based primarily on the data that is available at this higher level and on the ultimate use of the model.

At this higher level, the circular interconnections between the individual submodels mean that most of the data is generated from within the model during model operation. Only a few key inputs are provided, and the model simulates the remaining data. This is a desirable characteristics for making predictions for the future - the more aggregate and abstract the inputs are the more confidence we can have in their future values. But this is a less than ideal situation for estimating parameters - the model behaves as an aggregate model at the highest level even if its individual submodels work with disaggregate data. Hence in urban systems modelling the parameters to be estimated at the highest level are often those that strongly influence certain aggregate properties.

The purpose of the model must be taken into account at this point. The degree to which various aggregate outputs will be used in policy analysis dictates how important it is that the model should be accurate with regard to those outputs. The sensitivities of those outputs to various policy variables are another important aggregate characteristic of the model. Parameters that strongly influence these sensitivities are good candidates for adjustment when the highest level of estimation is performed.

Many of these parameters may have initial values already estimated within the individual submodels. But there are likely to be certain parameters that cannot be estimated except at this highest level - such as those that control the degree of interaction between submodels. Other parameters may have initial values estimated from the lower level models, but a lack of confidence in these initial values dictates that they can be better estimated using the entire model system.

8.2.4.1 Advantages and Disadvantages

Sequential estimation has many of the same advantages of simultaneous estimation. Some of the most important advantages are:

- various extra data can be used when estimating the lower level models, but the highest level of estimation can ignore these extra data,
- observed data is not required for all data flows because previously calibrated submodels can provide synthetic data, and
- the entire modelling system is adjusted in a systematic way to match observed data.

Sequential estimation is less accurate than simultaneous estimation (Brownstone and Small, 1989), and the error estimates on parameter values are biased (Amemiya, 1974). An important limitation appears when a parameter is shared between submodels (the α's in figure 15) or exists in one submodel but needs to be further adjusted at the highest level. The final value of such a parameter will be determined only by the last estimation procedure; the information on the parameter from earlier estimations will be discarded.

8.2.5 Bayesian Sequential Estimation

Bayesian parameter estimation theory allows for a function, called a prior density function, to specify what is already known about certain parameter values. This allows prior knowledge to be included in a parameter estimation process. The prior knowledge is often theoretical knowledge regarding the range of acceptable parameter values, but it can also include information from previous parameter estimation exercises.

In particular, the parameter estimates and confidence limits from the estimation of the parameters within individual submodels could be used when estimating the parameters at the highest level of the modelling system. This is a relatively untested method for urban systems models, but it would allow for a more theoretically pure way of estimating shared parameters without going to the complexity of full simultaneous estimation.

In the nested logit example of figure 13, the estimation of the mode choice submodel will provide estimates of the variance and covariance of the parameters. This information could be used to partially constrain any parameters that also occur in the location choice submodel, so that the final estimated value of the shared parameters are informed by all the available data.

8.3 Sacramento

The parameters in the MEPLAN land use and transport interaction model of the Sacramento, California region were estimated using a sequential approach. The submodels and some of the extra models are shown in figure 16.

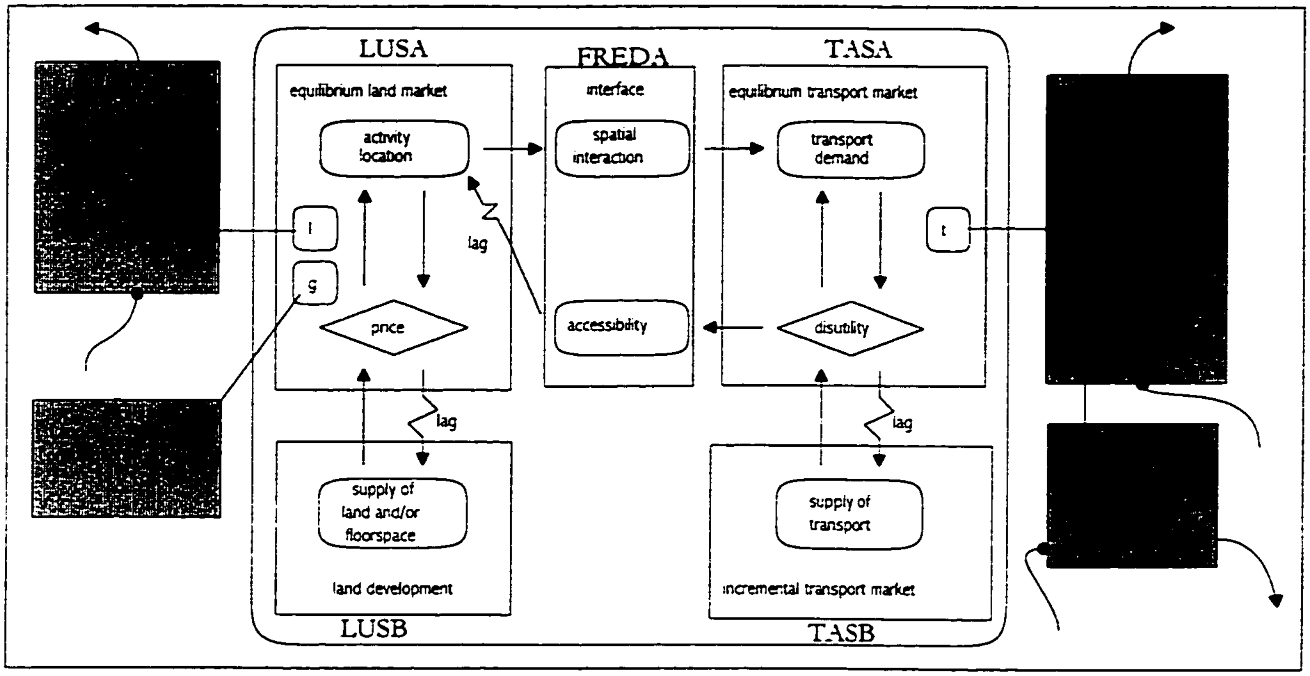


Figure 16: The submodels of the MEPLAN model of Sacramento and various extra models used to inform parameters

The extra models are shown along the left and right sides of figure 16. The parameters in the input/output economic model of Sacramento were estimated based on observed data on the relationships in the economy. These data were not considered directly in MEPLAN; instead various parameters (shown as λ) were taken from the input/output model.

The land consumption model is a representation of the tendency to consume less land as land prices rise. These relationships between density and price were explored in an extra model outside of MEPLAN so that "extra data" concerning the quantities of land developed in different zones could be considered. This "extra data" could, in fact, have been considered within the MEPLAN model, but it simplified the procedure to consider them in an extra model and to transfer the resulting parameters (γ) directly into MEPLAN.

The SACMET 4 step model was an existing transportation model that had been calibrated based on various aggregate data, with some of its parameters (τ) in turn coming from other models, including a disaggregate mode choice model (DKS Associates, 1994). The disaggregate data were "extra" in the sense that they could not have been directly used in MEPLAN's aggregate formulation. Many of the parameters in these models were used directly in the MEPLAN model.

The parameters in LUSB, TASB and FREDA were treated separately as well, although many of these parameters followed directly from assumptions or theory and were not "estimated" in a statistical sense.

It was not possible to consider LUSA model separately from the TASA model, however. The "spatial interaction" data at the centre of the top of figure 16 consists of detailed tables describing how much interaction occurs between different amounts of economic activity of different types in different zones. Observed data at the required level of detail were not available, hence TASA could not be run independently of LUSA. The accessibility numbers at the centre of figure 16 were not available either, so LUSA could not be run independently of TASA. Hence most of the parameters in both LUSA and TASA were estimated in the overall estimation process.

8.3.1 Overall Parameters

The values of the Mode Specific Constants (in TASA) should be similar to those in the SACMET model, but as "catch all" terms from different types of models the degree of similarity might be small. Further, these values control the key mode share predictions. Decision makers and the general public are accustomed to looking at mode shares when evaluating various future scenarios (chapter 6), so these parameters control an important output from a policy analysis perspective and were estimated so that the modelling system's prediction of mode share in the base year matched observed values.

Mode choice dispersion parameters (in TASA) should also be similar to those from the SACMET model, but the modal share proportions for each zone-pair provided data to reliably estimate these parameters within the overall model estimation.

These parameters provide an insight into how Bayesian sequential estimation might provide an advantage over sequential estimation. In Bayesian estimation, the values of these parameters from the SACMET model could be included in the likelihood function for the overall estimation, but the prior density function would indicate that the values from the SACMET model are not to be considered accurate.

In LUSA the spatial allocation model's dispersion parameters needed to be estimated in the overall estimation because there were no theories or assumptions or extra models or extra data that could be used to inform their values. Further, they influenced the amount of "spatial interaction" between zones, which influences the amount of travel, and the total amount of travel was an important policy output with observed data for the base year. Estimating these parameters at the highest level allowed for a good match with the observed data.

8.3.2 Objective Function

In chapter 9 the overall calibration will be described. Briefly, an objective function was created which considered the match between modelled and observed values for:

- various indications of the proportion of travel occurring by different modes,
- the spatial distribution of trips in the form of two Origin-Destination matrices, one for private automobile trips and another for trips by other modes, and
- the "trip length distributions" showing, for automobile trips, the proportion of trips that are of different lengths.

These data were used partially because they were measurements that were available, but in an urban system there is an almost infinite amount of data that could be collected. The data for estimation must be chosen based on the cost of obtaining the data and its relevance to the modelling exercise. These data in particular were chosen because they corresponded to model outputs that were both important and influenced by the parameters to be estimated.

The initial objective function was a quasi-likelihood Chi-Squared function formed by estimating the size of the variance of each of these measures, and assuming independent normal distributions. The resulting function to be minimized is the sum of squares of the differences between the modelled values and the measured values - a "weighted least squares" approach. The objective function would be a true "likelihood function" if the errors are independent measurement errors in the observed values. As discussed above (section 8.2.1), the greater cause of the errors is inherent randomness in human behaviour or inaccuracies in the model's equations, so it is more appropriate to call this function a "quasi-likelihood function" or just an "objective function".

The relative independence of different submodels and the degrees of freedom within individual models meant that certain parameter values could not be estimated properly with the data that was available. (For example, no data were available describing the proportion of work trips made by bicycle, so the mode specific constant for bicycle for work trips could only have been estimated through its complex secondary effects.) This is common in parameter estimation, as described in Press *et al* (1988) pg. 534:

> There is a certain mathematical irony in the fact that least-squares problems are *both* overdetermined (number of data points greater than number of parameters) *and* underdetermined (ambiguous combinations of parameters exist).

Press *et al* suggest using a Singular Value Decomposition approach to solve for the step during the quasi-Newton optimization procedure. This technique can be adjusted to avoid random large multiples of correlated adjustments to the parameters. However if these "ambiguous combinations of parameters" can be identified in advance there are at least two alternatives to the techniques of Singular Value Decomposition:

- The ambiguities can be eliminated in advance by pre-specifying an exact relationship between the parameters (as was done for the bicycle mode specific constant, where one mode constant was estimated for all trip types), or
- Prior knowledge (from theory or from previous experiments) about the relationships between the parameters can be included in the objective function in a Bayesian prior density function.

The initial quasi-likelihood objective function led to model parameters that were less than ideal, in the sense that they did not give enough accuracy in the key mode-choice outputs that would be viewed by policy makers. In particular, the parameter estimation was being driven largely by the goodness of fit in the origin-destination matrices because of the shear number of data points in the origin-destination matrices. This led to a sacrifice of goodness-of-fit in the mode choice values.

It would have been possible to continue to use the quasi-likelihood approach by finding more data on mode choice, but given the almost infinite amount of data available in an urban system it seems that selecting the amount of data in each category specifically to manipulate the statistics is itself inconsistent with the statistical theory. Instead, the weightings assigned to the different targets were adjusted so that the resulting model would predict the most important variables with suitable accuracy.

8.3.3 Overall Estimation Procedure

Chapter 9 describes the procedure used to optimize the objective function. Initially, the estimation of overall parameters was done manually, where various adjustments were made to certain parameters to attempt to improve the overall fit as measured by the objective function. As the relationships between various parameters and outputs became clear certain parameters were considered together, by finding the linear combination of the parameters that improved several targets at once.

Eventually, an automated heuristic search method (The Levenberg-Marquardt method, Marquardt, 1963) was used to further adjust the parameters to improve the goodness of fit. This made the process faster, more transparent, ensured repeatability in the final results and provided numerical matrixes indicating the patterns in lack-of-fit that could guide future decisions to change the model design as part of the larger calibration process.

MEPLAN already incorporates many iterative procedures, some that are necessary to calculate the model's outputs (i.e. to solve equation 19, the reduced form of the model) and others that estimate zone specific constant parameters. These iterations within the iterations of the heuristic search are time consuming and produce numerical inaccuracies that make the search more difficult than necessary. Some of these iterations could be disabled by estimating the zone specific constants within the overall estimation. The others could be eliminated by reformulating the likelihood function to be in terms of the structural form of the model. The likelihood function would be a direct function of the residuals in each of the model's *structural equations* (equation 20) instead of each of the model's outputs (equation 19). As mentioned above, this would be more theoretically appealing since the measurement errors in most of the data are considered small compared to the structural errors in the model's equations. However the MEPLAN software package cannot calculate these residuals — other software would need to be constructed.

8.3.4 Advantages of Sequential Estimation for Sacramento MEPLAN

The use of the sequential technique, where some parameters were taken from extra models that considered extra data and other parameters were established in an overall estimation, was very successful. The final model was adjusted to match known aggregate quantities, making "validation" unnecessary. Thus the entire modelling system is known to perform appropriately. Yet the model is constructed out of submodels that were considered individually, allowing a larger amount of data to influence the final parameter set. Certain behavioural parameters (e.g. mode choice parameters) were estimated in extra models that were able to discern the motivations of behaviour better than an aggregate model like MEPLAN could.

The "limited view" approach, where only the entire modelling system is investigated, might have produced a model with a good match to observed aggregate data. Yet the predictive ability and the behavioural nature of the model would have been called into question because various internal data flows might not have matched available observed data, and various behavioural parameters may have had counterintuitive values.

The piece wise approach, where individual submodels are calibrated, could not have been accomplished in the Sacramento MEPLAN model because the data needed to link the transport model to the land use model was not available. Even if such data were available, there would have been no guarantee that the final modelling system would have matched observed aggregate data when the pieces were connected.

The simultaneous approach was unrealistic because of the complexity of the various models and because some of the extra models were created by other organizations not involved in the MEPLAN modelling project.

The Bayesian sequential approach would have had most of the advantages of the sequential approach, yet more information could have informed some of the parameters. For example, the relationships between mode specific constants for bicycle travel from the disaggregate mode choice model would have partially constrained the bicycle mode specific constants in the overall estimation, and they would not need to have been constrained "artificially". Other parameters that were "fixed" in the final estimation could have been given a narrow prior density function instead if the Bayesian approach was used. This would have allowed them to vary if they had a strong influence in the overall performance of the final model, yet they wouldn't float around inappropriately because of complex secondary relationships. This is theoretically appealing, yet the nature of the current software implementation of MEPLAN is such that that the amount of required computing time would be very unreasonable if too many parameters are allowed to vary during the final estimation.

8.4 Implications

This chapter has described the application of certain statistical theories regarding parameter estimation in non-linear models to urban systems models. Most importantly, it has described some of the practical methods that are often used in estimating such models, and described the theory, the advantages and disadvantages of each method. The intricate details of statistical theory have been avoided in order to give practitioners a general understanding of the reasons why various methods might be used and to pique their interest in alternative techniques. Further details on the statistical methods are available in Press *et al* (1988) or in Bates and Watts (1988). An overview of using computer simulations as experiments is in Naylor (1971). A fairly complete and theoretically detailed exposition that does not presuppose a sophisticated statistical background is given in Bard (1974).

In most transport modelling research the emphasis has been on model theory, model operation or on detailed methods for estimating the parameters of simple submodels. Parameter estimation at the highest level has received very little attention. Yet in the practical development of large scale models for policy analysis the overall estimation may be the most important task. Practitioners develop submodels with rigorous statistical estimation techniques, then combine these submodels into larger frameworks without any attention to the errors that this may introduce. They then try to "validate" the overall model by comparing various predictions with known aggregate quantities with little guidance as to how this is best done. If the model does not meet expectations during this validation the model may be further adjusted, but this overall adjustment is seldom done in a theoretically defensible, objective manner.

This lack of emphasis on overall estimation may have been appropriate in the past. For instance, Anas (1987) argues that large scale urban models should conform to strong economic theories, implying perhaps that the development of the theories and methods of individual submodels needed to occur before the theories and methods of overall parameter estimation. As well, the computing power of a decade ago may have made overall parameter estimation practically impossible. Regardless, it seems that now is the time to shift our emphasis to the consideration of this important issue.

The TRANSIMS project in the US provides evidence that this shift is not yet happening. The TRANSIMS project has developed sophisticated activity simulators, population synthesizers, economic simulators and very ambitious traffic assignment methods. Yet making the full model system accurate for an individual region appears to have received scant attention to date.

In spite of the investigation of estimation theory in the Sacramento model, *ad-hoc* weights were eventually used in the overall estimation process. Unless the purpose and emphasis of the model can be specified mathematically and a pure likelihood equation can be written the overall calibration will still need some *ad-hoc* adjustments. Minimizing such intuitive manual intervention is a worthy goal, but eliminating such intervention may not be possible or even desirable.

One way to further reduce intuitive manual intervention in the final estimation process is to use the confidence information from previous estimations as a prior density function included in the likelihood function. This Bayesian sequential estimation process should be pursued in further work.

Large scale urban models emphasize the calculation of model outputs. That is, the emphasis is on the reduced form of the model. Many of the errors in urban systems modelling are structural errors, and different software implementations are required to calculate likelihood functions based on structural errors. These structural error calculations should be computationally simpler, suggesting that the resource costs of estimating model parameters could be lessened by adopting a more theoretically pure likelihood function. This is a research topic that should be pursued in greater detail.

9 Semiautomatic Calibration

9.1 Chapter overview

The parameters of the Sacramento MEPLAN model that was developed, described in Part 2 of this dissertation, were estimated following the sequential procedure described in the previous chapter. The overall estimation of the highest level parameters was done using a manual search procedure. The manual search procedure was fairly systematic, which lead to insights into the appropriateness of different automated procedures.

This chapter describes both the manual procedure (that lead to the model described in Part 2) and the automated procedure that was eventually developed based on the knowledge gained from the manual procedure.

9.2 Manual Search procedure

The highest level parameters estimated in the estimation procedure were:

- mode specific constants
- dispersion parameters for each factor for production allocation
- mode choice dispersion parameters
- zone specific constants for production allocation

These highest level parameters were estimated by comparing a selection of model outputs with observed data ('targets'). The merit function used to measure the goodness of fit was a weighted sum of squares of differences:

$$S(x) = \sum_{t=1}^{n}\left(\frac{y_m^t(x) - y^t}{\sigma_t}\right)^2 \qquad (21)$$

where:

S	= the merit measure, with lower values corresponding to a better goodness-of-fit
y^t	= the measured value of target t
$\mathbf{x}$	= the vector of parameters
$y_m^t(x)$	= the modelled value of target t when the parameters have the values $\mathbf{x}$

σ_t = a measure of the "unimportance" of target t. Always positive, with a smaller value indicating that it is less important to match target t.

The manual search procedure involved investigating each target either individually or together with closely related targets, and determining which parameter would most strongly influence that target. The chosen parameter was adjusted to improve the goodness-of-fit of the chosen target; and the overall goodness-of-fit from equation 21 was monitored. The parameter was adjusted to minimize the overall goodness-of-fit. Mathematically, this is a line search along a single axis to find the minimum of the objective function along that line.

After each such line search it sometimes occurred that the goodness of fit of the chosen target was not "good enough" given the purpose and use of the model. In this situation the details of each target needed to be investigated to determine which other targets were pulling the parameter in the other direction and prohibiting the chosen target from achieving a minimum. This information about how other targets influenced a single target through a single parameter was found to be vitally important in the search for a good model. Three examples will serve to illustrate this.

9.2.1 *Davis influence on mode choice targets*

First, the attempts to match mode shares exactly by adjusting mode specific constants was thwarted by the goodness of fit in the OD matrix. When this was investigated, the characteristics of travel within a single zone were found to be strongly influencing the overall goodness of fit. The single zone was the city of Davis, where the University of California at Davis is located, and where a large number of students and staff use non-motorised transport in a cultural and physical setting that encourages such travel. As well, many of the students at Davis are temporary residents with unique commercial and social needs, and substantial agglomeration economies are available by clustering these students spatially within the City of Davis.

Davis is only one small zone in the MEPLAN model of the Sacramento Region, however, and it was more important to accurately model the remainder of the region than to model the City of Davis. Thus, firstly the weight on the Origin Destination matrix was reduced to make it less important in the overall goodness-of-fit. This did not seem to solve the problem, however, so eventually the student and staff of Davis were represented separately in the model.

This shows how the relationships between different targets, the parameters and the target weights suggested first-off a change in the weights, but eventually led to a change in the model design.

9.2.2 *Splitting of the Office-Services factor*

The search for goodness of fit in the trip length distribution for "home-other" trips involved changing the location choice dispersion parameter for a variety of industrial categories that could have been the destination for such trips. These were agriculture, manufacturing, office/services, health and government. It was possible to make the modelled trip length distribution match the observed values fairly closely by adjusting this parameter, but the overall goodness of fit suffered. An investigation of this should that the "other-other" trip length distribution was becoming much too short when this parameter was large enough to match the "home-other" trip length distribution. Thus, these two different sets of targets were "pulling" on the parameter in different directions. This was not surprising, as the design of the model had the interrelationships between these factors generating the "other-other" trips, while the consumption of these factors by households generated the "home-other" trips.

An analysis of this showed that the largest generator of "home-other" trips was the consumption of office/services by households. The location parameter for this factor was adjusted separately, which improved the problem somewhat but it was still not possible to obtain a good match to both trip length distributions. It was found that the largest generator of "other-other" trips was the consumption of office/services by other industry. Thus if "office/services" was especially sensitive to transport costs then the "home-other" trips would be short enough but the "other-other" trips would be too short. One the other hand, if "office/services" was not that sensitive to transport costs then the "other-other" trips would be long enough but the "home-other" trips would be too long.

A possible explanation for this is that "office/services" supplied to households is a more homogeneous commodity and hence there is little need to travel long distances to obtain a unique product, while "offices/services" provided between industry represents a broader category of interactions. Note, though, that the "other-other" trips are shorter overall than the "home-other" trips, because the spatial clustering of industrial activity in certain zones allows a wide range of access to different services without requiring long travel. Thus even though long travel is more likely for other-other trips than for home-other trips *all else being equal*; all else is *not* equal in the spatial arrangements of different activity in the Sacramento region. The full interactions represented in the modelling system are necessary to uncover this; a piece-wise estimation without an overall estimation might miss this important aspect of the situation in Sacramento.

This realization led to the splitting of the "office/services" factor into two separate factors, one representing the supply of services to households and the other representing the supply of services between industry.

This shows how the relationships between different targets, the parameters and the target weights in the overall calibration led to a change in the model design.

9.2.3 Simultaneous search in several parameters

The mode choice dispersion parameter controls the sensitivity of mode choice to the attractiveness of each mode. In the zonal aggregate level represented by MEPLAN the proportion of people choosing different modes between different zone pairs provides an indication of the number of people choosing different modes under different conditions. Thus the Origin-Destination trip tables by different modes contained information that could establish a value for this parameter.

This parameter was adjusted to improve the goodness-of-fit in the Origin-Destination trip tables, but this would make the goodness of fit in the overall mode split much worse. Since the overall mode split was the more important output for policy analysis, this parameter was not changed very much at all. When the scenarios were initially ran, the sensitivity of mode choice to policy inputs seemed exceptionally low. This parameter was revisited, and the mathematics of logit choice models was investigated, until it was discovered that the product of the mode specific constants and the dispersion parameter could be held constant while adjusting the dispersion parameter. This allowed the mode share predictions to remain fairly stable while calibrating to the mode choice sensitivity. Thus a search along a line that was not parallel to any of the parameter axes was necessary to improve the overall goodness-of-fit while improving the mode choice sensitivity.

The automated search algorithm described in the next section is able to mathematically identify such search directions and hence improve many targets simultaneously.

9.3 Automatic Search Procedure

If the errors in the model are all assumed to be independent measurement errors in the y^t and if the σ_t in equation 21 are taken to be the variances in this measurement error, then equation 21 is the "chi square" and it is the (variable part of the negative) log of a likelihood function (Bard, 1974). However, the previous chapter showed that these assumptions are unrealistic for large scale urban models. Even if the assumptions were valid, the parameter estimation process needs to take into account the ultimate purpose of the model. The σ_t will be adjusted in a manual, intuitive way so that the final model can reproduce those targets that are important for the purpose of the model. Equation 21 should therefore be called a "goodness-of-fit" function, a "merit function" or perhaps a "quasi-likelihood function" but not a "likelihood function".

Equation 21 can be rewritten in matrix notation:

$$S(x) = [\Lambda \cdot (y - y_m(x))]^T \cdot [\Lambda \cdot (y - y_m(x))] \tag{22}$$

where:

$$\Lambda = \begin{matrix} \frac{1}{\sigma_1} & 0 & \cdots & 0 \\ 0 & \frac{1}{\sigma_2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \frac{1}{\sigma_n} \end{matrix}$$, a diagonal matrix of the weights

y = a vector of observed target values,

$\mathbf{y}_m(\mathbf{x})$ = a vector of modelled values at parameter values **x**

The necessary conditions for a minimum of equation 22 can be found by taking its derivative with respect to **x** and setting it to zero, giving the "normal equations":

$$\mathbf{0} = \boldsymbol{F}^T \cdot \Lambda^2 \cdot (y - y_m(x) \tag{23}$$

where:

$$\mathbf{F} = \begin{matrix} \frac{\partial y_m^1}{\partial x_1} & \cdots & \frac{\partial y_m^1}{\partial x_m} \\ \frac{\partial y_m^2}{\partial x_1} & \cdots & \frac{\partial y_m^2}{\partial x_m} \\ \frac{\partial y_m^3}{\partial x_1} & \cdots & \frac{\partial y_m^3}{\partial x_m} \\ \vdots & \ddots & \vdots \\ \frac{\partial y_m^n}{\partial x_1} & \cdots & \frac{\partial y_m^n}{\partial x_m} \end{matrix}$$, a matrix of the partial derivatives of the various elements of **y** with respect to the various elements of **x**

To use equation 23 in the search for the best fit, it is necessary to find the matrix of partial derivatives **F**. However the relationship between the modelled output variables $\mathbf{y}_m$ and the parameters $\mathbf{x}$ is usually complex, as it involves solving the reduced form of the model (equation 19 in chapter 8). Thus these values were approximated through numerical differentiation. Each parameter was perturbed slightly in one direction, and the changes in the various modelled output variables gave an indication of the values of one column of **F**. Hence the number of model runs required to find an approximation of **F** was equal to the number of parameters investigated at this highest level.

The numerical accuracy in the MEPLAN implementation led to problems in finding an appropriate size for the perturbations Δx_j in each of the *m* parameters x_j . If the perturbation was too small, then the numerical inaccuracies in calculating each of the y_m^i could be larger than the change in y_m^i , leading to a completely inaccurate approximation of the partial derivative $\frac{\partial y_m^i}{\partial x_j}$. However, if the perturbation was too large, the secant approximation to the tangent becomes less accurate as well. A perturbation needed to be large enough to overcome the numerical inaccuracy in *each* of the *n* targets y_m^i while still being as small as possible.

The partial derivative matrix **F** can be used in a Taylor Theorem approximation for the modelling system:

$$\boldsymbol{y}_m(x) \approx \boldsymbol{y}_m(\boldsymbol{x}_0) + \boldsymbol{F} \cdot \Delta \boldsymbol{x} \qquad (24)$$

$$\Delta x = \boldsymbol{x} - \boldsymbol{x}_0$$

which approximates how the entire modelling system will respond to small changes in the parameter values.

Equation 24 can be inserted into the normal equations (equation 23) and solved for Δx , the changes in the parameter values that would minimize equation 22 if equation 24 were completely accurate. This is:

$$\Delta \boldsymbol{x} = [\boldsymbol{F}^T \cdot \Lambda^2 \cdot \boldsymbol{F}]^{-1} \cdot \boldsymbol{F}^T \cdot \Lambda^2 \cdot (\boldsymbol{y} - \boldsymbol{y}_m(\boldsymbol{x}_0)) \qquad (25)$$

Repeatedly calculating and applying Δx in this way is called Newton's method and gives Order 2 local convergence (Bard, 1974). But too large of a Δx can overwhelm the approximation implied of equation 24, and lead to divergence. Further, equation 23 might be a maximum point or a saddle point, instead of a minimum point. A method must be used to scale back the parameter changes, and ensure a search in a downward direction.

A sufficiently small step in the steepest descent direction is guaranteed to improve the goodness of fit:

$$\Delta \boldsymbol{x} = d \cdot \boldsymbol{F}^T \cdot \Lambda^2 \cdot (y - y_m(\boldsymbol{x}_0)) \tag{26}$$

where:

d = is a constant adjusting the size of the step.

If the steepest descent direction is used to find a shorter step when Newton's method begins to diverge, then the overall search is guaranteed to converge. The similarities between equations 25 and 26 suggest that d should never be so large as any of the diagonals in the matrix $[\boldsymbol{F}^T \cdot \Lambda^2 \cdot \boldsymbol{F}]^{-1}$ — if a step this large is desirable then it is better to use Newton's method directly. The steepest descent method is to be used only to find smaller step sizes when Newton's method begins to diverge.

The *Levenberg-Marquardt Method* (Bard, 1974; Press *et al*, 1988) can be used to smoothly move between the two extremes of equation 25 and equation 26. This method involves augmenting the diagonal of $\boldsymbol{F}^T \cdot \Lambda^2 \cdot \boldsymbol{F}$ by a fixed multiplier $1 + \lambda$. When λ is near zero the Newton step will be used. When λ becomes large a small step will be taken in the steepest descent direction. In between, the step will be a "scaled back" Newton step — a smaller distance than the Newton step and in a direction closer to the steepest descent direction.

The algorithm works by taking a step in the calculated direction. If the step leads to a better goodness of fit, then λ is reduced by a fixed multiple and the next step proceeds from the new position. If the step leads to a worse goodness of fit, then the step is abandoned, λ is increased by a fixed multiple and a new step is calculated.

This algorithm required considerable modifications to deal with the numerical inaccuracies introduced when solving for the models output variables y. It is never certain whether the goodness of fit is actually increasing or decreasing, because it is always possible that the numerical inaccuracies are overwhelming the changes in the targets. For the most part, the modifications were changes to the "search parameters", which are numbers that control the search process. However it was also necessary to discard (and recalculate) the information in the **F** matrix and in the modelled target values $\mathbf{y}_m$ if the algorithm had scaled back its step size a few times and was still not able to improve the goodness of fit.

9.3.1 Constraints and zone specific constants

MEPLAN has a constraint process that will adjust the prices and/or disutility of a factor in a zone so that the amount of that factor in the zone matches a predetermined value for any one year. In future years this procedure is used to represent a short-run equilibrium for some factors (e.g. land), where prices do not match input costs but are set through this iterative process that matches demand to a fixed amount of supply.

This same process is commonly used in the calibration year to constrain the amount of activity in zones to match the observed amounts, so that in the base year urban form and spatial distribution matches observed values exactly. The adjustments in the disutilities needed to match these constraints in the calibration year are taken as the "zone specific constants" representing all those things that contribute to the attractiveness of a zone for a specific activity that are not explicitly included in the model.

Currently, this constraint process is considered part of the model's equations, and the zone specific constants are not considered "parameters" in the model. There are a number of other ways to consider this constraint process:

- The zone specific constants, as "catch all" terms, represent information that could not be included in the mathematics of the model. There is a strong argument that these should be as small as possible, or at least contain the smallest amount of information. The size of these terms, or some measure of their spatial distribution and correlation, could be included in the objective function, so that the parameter estimation process can find the model that relies the least on these parameters. This would have the added benefit of exposing the zone specific constants in the estimation process, allowing them to be inspected by the model builder. Unrealistic zone specific constants can be a problem in MEPLAN models, as shown in chapter

7. Being able to find the relationships between excessive zone constants and other targets and parameters might be useful in pinning down the causes of problematic zone specific constants.

- The zone specific constants could be considered as parameters in the model, and the mismatch between modelled and observed activity amounts could be considered a target. Thus the model might not exactly match the amount of activity in a zone if that target was at odds with some other observed value. (If an exact match is desired, the weights could be increased for those targets, causing them to act as a penalty function.) The full intersection of factors and zones gives a very large number of targets and parameters in this method, leading to very large matrices in the search process. It would certainly be necessary to use symbolic derivatives in forming the relevant columns of the **F** matrix, and it might also be necessary to exploit the resulting structure of the matrix when solving the various linear systems in the search process. This process might work well if the number of zone specific constants was reduced *a priori* through knowledge of the local geography.
- The zone specific constants could be considered as parameters and the zonal amounts as targets (as above), but the Lagrange Multiplier method could be used to attempt to symbolically eliminate them from the numerical optimization procedure. This might be achievable if calculating the objective function did not rely on numerical procedures, such as if the objective function was expressed as errors in the structural form of the model (see chapter 8 and Bard, 1974, page 25.)

9.4 Information achieved at convergence

9.4.1 Information Matrix

When the model has converged, the information matrix is given by

$$[\boldsymbol{F}^T \cdot \Lambda \cdot \boldsymbol{F}]^{-1} \tag{27}$$

This matrix contains the error information on the parameters, showing how accurately each parameter is estimated, and the covariance between parameter values, from which the correlations can be calculated. The correlations are useful numbers that show how pairs of parameter values are related. The problem of "multiple correlations" is well known in parameter estimation, where there is not enough independent variation in the data to distinguish whether one parameter value or another should be adjusted. As long as the same multiple correlations are still present when using the model for predictions this does not cause too much trouble. But often the scenarios to be investigated are ones that specifically break these correlations. The correlation matrix can show which parameters are affected by the correlations, and if these parameters are behavioural it is possible to identify the associated weaknesses in the model's predictive ability.

This matrix, with its information on parameter accuracy and correlations, is often used to guide further data collection exercises. More data could be collected to "tighten up" the error on a specific parameter value. More orthogonal data, such as stated preference data, could be collected to reduce the correlations between parameter values.

9.4.2 *Weight Sensitivity Matrix*

Since the weights used in the parameter search process are subject to ad-hoc adjustment, the sensitivity of the model to changes in these weights should be established. Assume that the estimation process has converged at a minimum point, but then the weights, Λ , are changed slightly so that $\Lambda_1^2 = \Lambda^2 + \delta\Lambda^2$. The next step $\Delta \boldsymbol{x}$ is calculated based on equation 25:

$$\boldsymbol{F}^T \cdot \Lambda_1^2 \cdot (y - y_m(\boldsymbol{x}_0)) = \boldsymbol{F}^T \cdot \Lambda_1^2 \cdot \boldsymbol{F} \cdot \Delta \boldsymbol{x} \tag{28}$$

or

$$\boldsymbol{F}^T \cdot (\Lambda^2 + \delta\Lambda^2) \cdot (y - y_m(\boldsymbol{x}_0)) = \boldsymbol{F}^T \cdot (\Lambda^2 + \delta\Lambda^2) \cdot \boldsymbol{F} \cdot \Delta \boldsymbol{x} \tag{29}$$

or

$$\boldsymbol{F}^T \cdot \Lambda^2 \cdot (y - y_m(\boldsymbol{x}_0)) + \boldsymbol{F}^T \cdot \delta\Lambda^2 \cdot (y - y_m(\boldsymbol{x}_0)) = \boldsymbol{F}^T \cdot (\Lambda^2 + \delta\Lambda^2) \cdot \boldsymbol{F} \cdot \Delta \boldsymbol{x} \tag{30}$$

The first term on the left is zero at convergence (from equation 23), so in the limit as $\delta\Lambda$ approaches zero

$$\boldsymbol{F}^T \cdot \delta\Lambda^2 \cdot (y - y_m(\boldsymbol{x}_0)) = \boldsymbol{F}^T \cdot \Lambda^2 \cdot \boldsymbol{F} \cdot \Delta \boldsymbol{x} \tag{31}$$

and it is possible to solve for the $\Delta \boldsymbol{x}$ caused by a small change in the weights.

It is possible to create a matrix which summarizes all this sensitivity information. If the weight of only one target, i, is changed, then the change in the parameters will be

$$\frac{\partial \boldsymbol{x}}{\partial\left(\frac{1}{\sigma_i^2}\right)} = [\boldsymbol{F}^T \cdot \Lambda \cdot \boldsymbol{F}]^{-1} \cdot (\boldsymbol{f}_i)^T \cdot (y^i - y_m^i(\boldsymbol{x})) \cdot \sigma_i^2 \tag{32}$$

where:

$(f_i)^T$ = the ith row of F, transposed into a column.

Thus

$$X_c = [F^T \cdot \Lambda \cdot F]^{-1} \cdot F^T \cdot \text{diag}(y^i - y_m^i(x)) \tag{33}$$

gives a matrix showing how the parameter values will change in response to relative (proportional) changes in the vector of weights. We will call this the Weight Sensitivity Matrix or simply X_c . The author is unaware of any previous use of this matrix in urban systems modelling.

This matrix is quite large. Interactive methods for inspecting and using the matrix are given in section 9.8.1.

The corresponding change in all of the targets due to a proportional change in the ith weight can be found by taking a column of X_c and inserting it into equation 24:

$$y_m(x) - y_m(x_0) \approx +F \cdot [X_c^T]_i^T \tag{34}$$

where:

$[X_c^T]_i^T$ is simply notation for the ith column of X_c .

9.5 Goodness of fit for model used for policy analysis

9.5.1 *Individual Targets*

The targets used in the overall goodness-of-fit calculation were:

- Mode split between private auto and other modes for each type of trip;
- Overall mode split between bicycle, walk, transit, SOV and HOV;
- Travel time distributions for automobile travel split by trip purpose; and
- Trip distributions by O/D pair and by auto vs. other modes.

Tables 4 through 6 compare modelled (calibrated) and observed values for mode split in the initial hand calibration.

		Modelled	Observed
Public modes mode split	Walk	55.69%	54.92%
	Cycle	25.20%	25.89%
	Transit	19.11%	19.19%

Table 4: Modelled and observed mode split of non-auto modes - all flows (hand calibration)

	Work-Low	Work-Med	Work-High	Education	Services
observed	10.03%	8.53%	7.89%	40.63%	10.58%
modelled	10.23%	8.10%	7.51%	39.51%	9.33%

Table 5: Modelled and observed percent by non-auto modes - by flow (hand calibration)

	Work (Low, Med, High)	Education	Home-Other	Retail	Other-Other
observed	17.86%	90.60%	41.50%	53.00%	52.50%
modelled	18.60%	82.92%	42.04%	52.98%	52.45%

Table 6: Modelled and observed mode split - HOV as a percentage of auto trips, by flow (hand calibration)

Figures 17 through 21 compare the modelled and observed distributions of auto trips by different trip type and travel time for the hand calibration. The parameters that most closely control the trip length distributions are the dispersion parameters for the allocation of production. These are location elasticities with respect to travel disutilities - a central component of a land-use/transport model.

Figures 17 through 21 exhibit a common trend: in general the number of trips in the longest categories are somewhat higher than for the target data. In future work these could be improved through the use of variable trip generation rates, where smaller numbers of physical trips are necessary for the same economic relationships when travel times are longer.

9.5.2 *OD Matrix Goodness of Fit*

The OD matrix goodness of fit is calculated as

$$\sum_{O} \sum_{D} \left(\frac{t^{o}_{OD}}{T^{o}} - \frac{t^{m}_{OD}}{T^{m}} \right)^2 \qquad (35)$$

where:

t^o_{OD} = number of observed trips from origin O to destination D

t^m_{OD} = number of modelled trips from origin O to destination D

O = ranges across the set of origin zones

D = ranges across the set of destination zones

T^o = total number of observed trips

T^m = total number of modelled trips

9.6 Goodness of fit for model estimated using the automated search.

Tables 7 through 9 compare modelled (calibrated) and observed values for mode split in the automated search calibration.

		Modelled	Observed
Public modes mode split	Walk	54.93%	54.92%
	Cycle	26.05%	25.89%
	Transit	19.02%	19.19%

Table 7: Modelled and observed mode split - all flows

	Work-Low	Work-Med	Work-High	Education	Services
observed	10.03%	8.53%	7.89%	40.63%	10.58%
modelled	10.11%	8.64%	8.00%	40.61%	10.70%

Table 8: Modelled and observed mode split - non-auto percentage of trips, by flow

	Work (Low, Med, High)	Education	Home-Other	Retail	Other-Other
observed	17.86%	90.60%	41.50%	53.00%	52.50%
modelled	18.04%	90.50%	41.64%	53.15%	52.20%

Table 9: Modelled and observed mode split - HOV as a percentage of auto trips, by flow

Figures 17 through 21 show the modelled and observed distributions of auto trips by different trip type and travel time for the automated calibration.

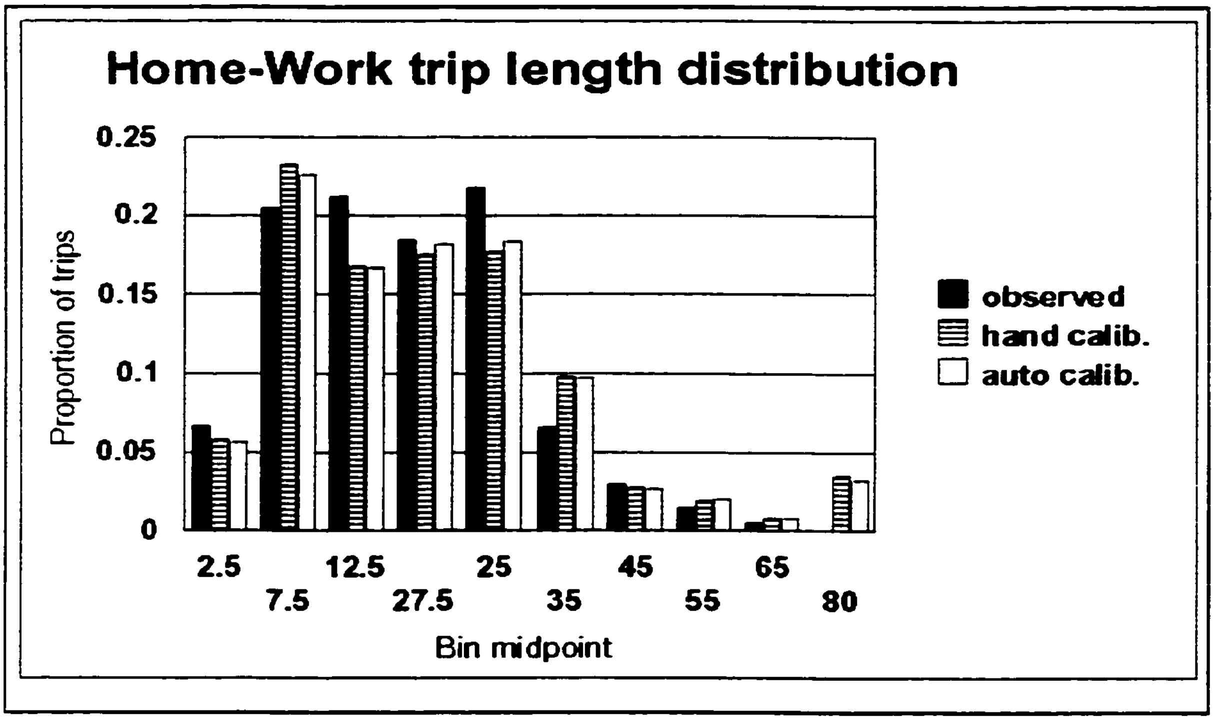


Figure 17: Observed and modelled Home-to-Work travel time distributions for automobile travel

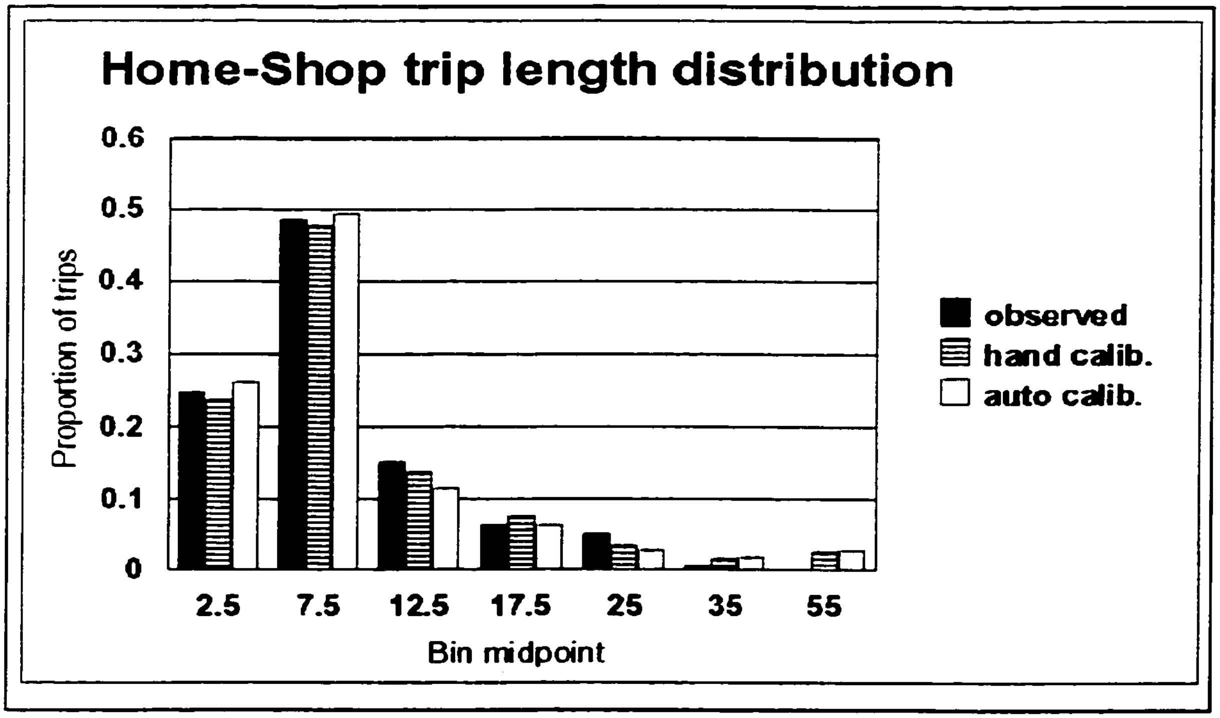


Figure 18: Observed and modelled Home-to-Shop travel time distributions for automobile travel

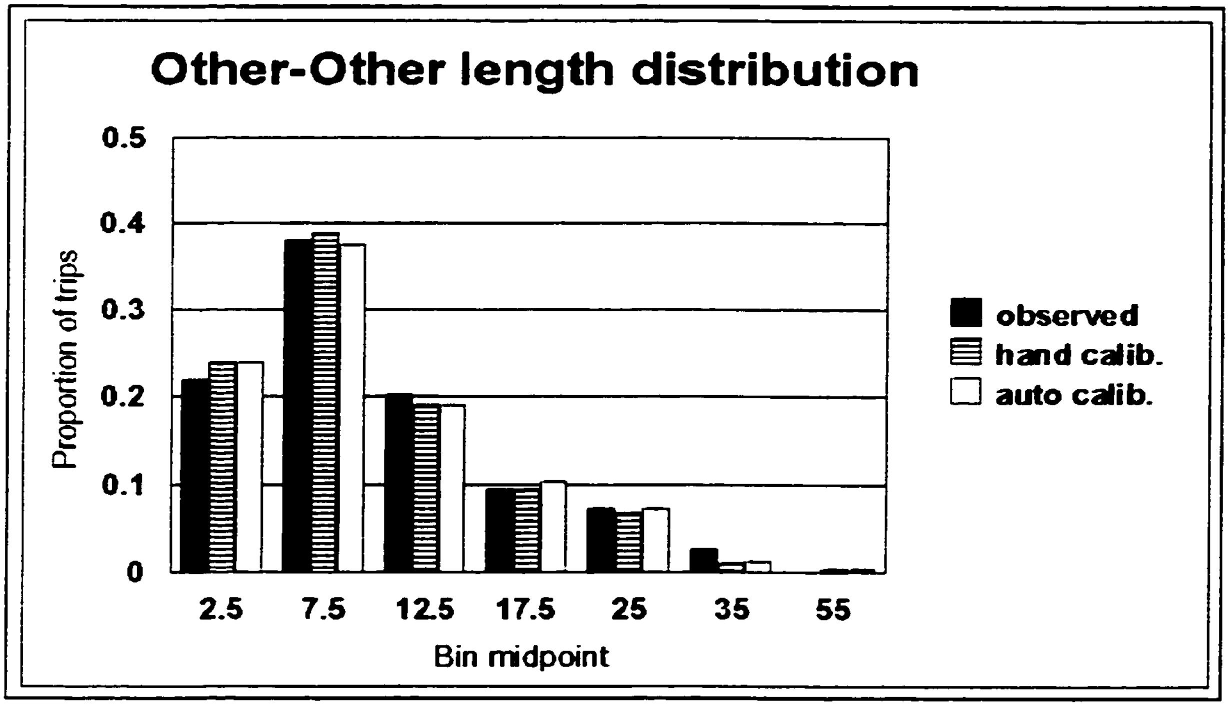


Figure 19: Observed and modelled Other-to-Other travel time distributions for automobile travel

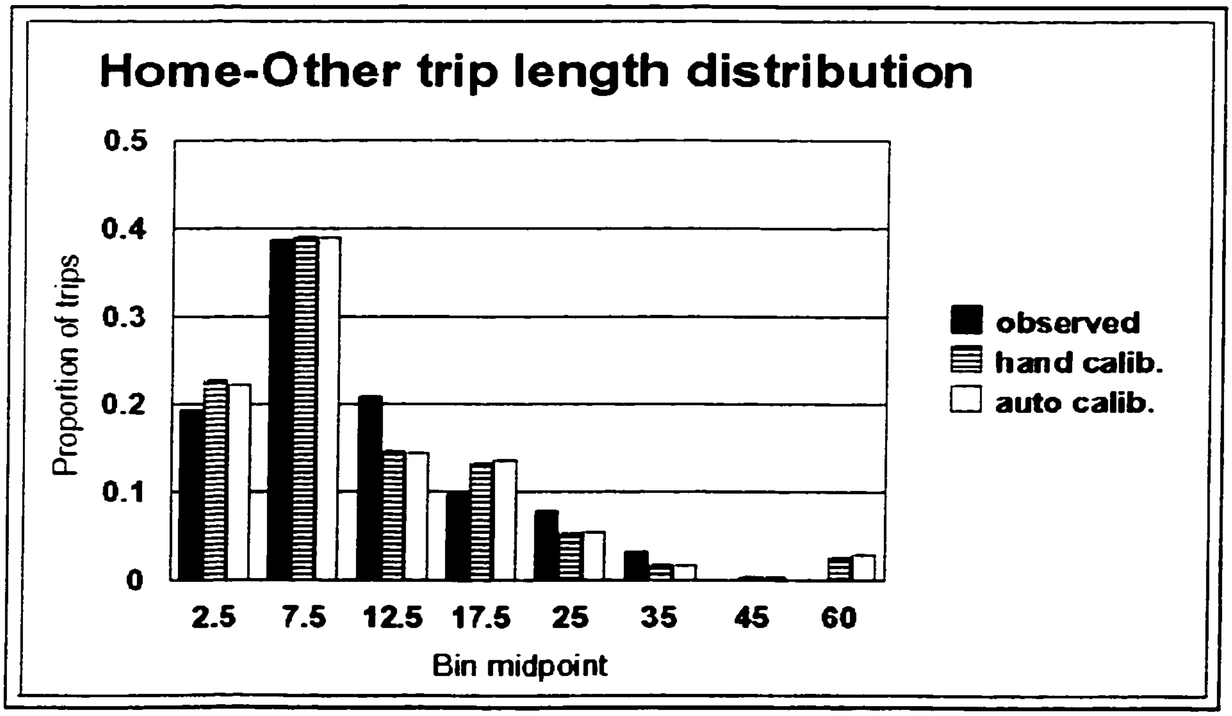


Figure 20: Observed and modelled Home-to-Other travel time distributions for automobile travel

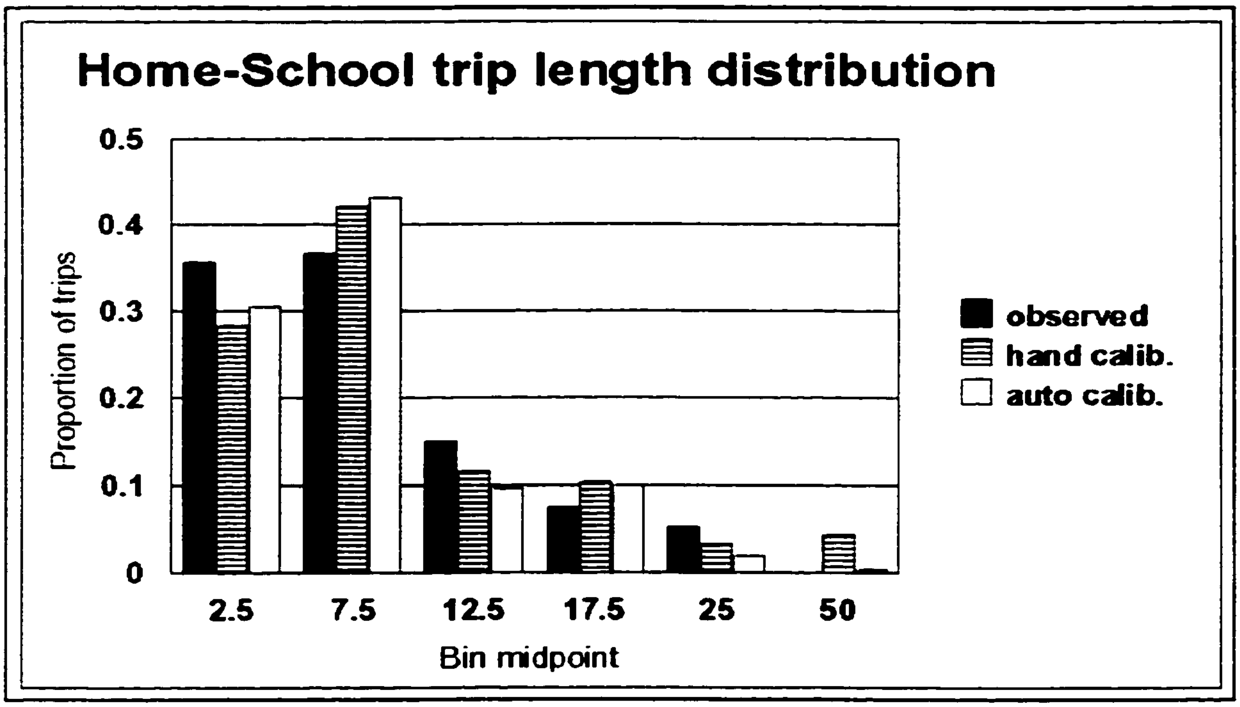


Figure 21: Observed and modelled Home-to-School travel time distributions for automobile travel

9.6.1 *Overall Goodness of Fit*

The goodness of fit function for the automated search process is the function in equation 22. The weights are as follows:

Target	Measured as	Shown	Number of targets	Weight, $\frac{1}{\sigma}$
Public vs. Private Mode Split	Portion of trips made by cycle, walk or transit (by flow)	Table 8	5	4.1
HOV vs SOV Mode Split	Portion of auto trips made by HOV (by flow)	Table 9	5	2.2
Walk, Transit use	Portion of all non-auto trips made by walking, transit	Table 7	2	5.77
Trip length distributions	Portion of all auto trips of a certain flow that fall between the travel time histogram bands	Figures 17 to 21	38	0.75
Public OD Matrix	Of all the non-auto trips in the region, the portion that occurs between each zone pair		3,249	2
Private OD Matrix	Of all the auto trips in the region, the portion that occurs between each zone pair		3,249	2

Table 10: weights used in automated search

The weights were initially based on the concept that each of the target types (i.e., each of the rows in table 10) should influence the final model approximately equally. Thus the weights were based on the average size of a measurement in the category and the number of data items in that category.

The simpler discussions of estimation theory suggest treating each data *item* equally, and the estimated parameters are those giving the model that is most likely to reproduce *all* of the data items. In this method the categories of data that necessarily contain more data items, such as OD matrices, would influence the final model more than the categories of data that contain less data items. A better approach for this project is to perform the estimation such that the model is just as likely to reproduce all the data items in one category as it is to reproduce all the data items in another category. This can be accomplished by scaling the weights by $\frac{1}{\sqrt{n}}$, where n is the number of data items in the category.

During the manual calibration an estimate of the variance in each output variable was made, and the weights for the weighted-least-squares merit function were calculated based on this variance and on the number of data items in each category. Throughout the calibration process it was found that the model was not reproducing the key outputs with sufficient accuracy. It was necessary to adjust upward the weights for the key outputs that would be viewed by policy makers. Thus, practical considerations regarding the purpose of the model were used together with statistical considerations to determine the final values of the weights.

The overall goodness-of-fit for the automatic calibration was 0.0590. For the hand calibration the overall goodness of fit was 0.0994. The automatic calibration achieved a better fit (lower values indicate a better fit) because it could consider all the cross-relationships together and could tirelessly search using an algorithm that has approximately order-2 convergence.

9.7 Calibration Software

The software used to run the MEPLAN model and search for the appropriate parameters was developed specifically for this dissertation. It was constructed using object oriented principles in the Java programming language.

Object-oriented programming works in terms of objects, so the software design description will proceed likewise. A `fixed width courier font` is used to identify software objects.

9.7.1 *Calibration Strategy*

A `CalibrationStrategy` is an object containing the `Parameters` that are to be adjusted, the merit function, and the procedure for finding the parameters that optimize the merit function. A generic `CalibrationStrategy` consists of a `Vector` of `Parameters`, a number of `Targets` organized in a tree structure for calculating the goodness of fit, a calibration method and a `ModelRunner`, defined below.

One particular `CalibrationStrategy` has been developed, the `LeastSquaresLinearized` strategy. This strategy implements the Levenberg-Marquardt method for performing weighted least squares estimation as described in the previous section.

9.7.2 *Model Runner*

A `ModelRunner` is an object that contains the code or procedures necessary to run the actual model that is being estimated. This is the procedural interface between the estimation method and the model itself. A `BatchSemaphoreRunner` has been developed, which will create an empty file (called a *start semaphore*) when it wants the model to run, and then wait until another file (the *stop semaphore*) is created by the model run itself. Another process watches for the existence of the start semaphore, then runs the model, then creates the other semaphore to tell the estimation process that the model run is finished.

For the Sacramento MEPLAN model, a dedicated machine was loaded with the DOS operating system and was networked together with the machine running the estimation program. The estimation program on one machine communicated with a DOS batch program on the other machine through these semaphores.

A `MeplanBatchRunner` is a further specialization of a `BatchSemaphoreRunner` that invalidates all the MEPLAN variables when the model is finished running, to tell the remaining sections of the software that the output files from the model run need to be re-parsed by the estimation software.

9.7.3 *Targets*

A `Target` is a Java interface encapsulating an observed target, the associated modelled value, the error between the observed and modelled values and the importance of the target in terms of its weight. There are many types of `Target`. A `Ratio` is a ratio between one `Target` and another `Target`, a useful construction that allows the estimation procedure to look at shares and relationships instead of absolute numbers. A `TripCount` is a target designed specifically for working with MEPLAN. It looks into the MEPLAN output file to find the amount of travel by specific modes and flows.

The detailed information on trips is specified by zone pair in the MEPLAN output file TAM.DAT. Two different `Targets` can be used to inspect this detailed information. A `TAMAggregateByZones` can sum up the information in TAM according to groups of zone pairs. A `TAMAggregateByData` iterates through all the zone pairs in TAM.DAT, summing up those that meet a certain criteria. Targets of type `TAMAggregateByZones` are used for the OD Matrix targets, while targets of type `TAMAggregateByData` are used for the trip length distribution targets.

9.7.4 *Dialogable Target Group*

A `DialogableTargetGroup` is an interface used to describe a group of targets. An `ODMatrix` is a type of `DialogableTargetGroup` that contains all the information on zone pair data flows. A `TripLengthDistribution` is a `DialogableTargetGroup` that contains a histogram of information on the length of trips of a specific type by a specific mode. A DialogableTargetGroup knows which View-Controller object (see below) can present all of the targets in the group to the user for modification using the keyboard and mouse.

9.7.5 *Parameter*

A `Parameter` is a class that represents the values that are to be adjusted in the model in the attempt to match the `Targets` and improve the goodness of fit. Three types of `Parameters` are used for Sacramento MEPLAN. A `ModeSpecificConstantM` is a mode specific constant in a MEPLAN UTP.DAT file. A `ModeSpecificConstantDifferenceM` is a way of representing the notion that the difference between assorted pairs of mode specific constants should be equal. The difference is the parameter that is adjusted, and in the UTP.DAT file the second constant in each pair is modified appropriately. A `MeplanUFileParameter2` is a representation of almost any data item that occurs one or more times in a "group" (see below) of a single MEPLAN input file. For Sacramento, this is used to represent the mode choice dispersion parameter and all of the location choice dispersion parameters.

9.7.6 *View-Controllers*

Many of these objects need to be manipulated on the computer screen using a mouse and keyboard. The *Model-View-Controller* paradigm central to Java user interface design was adopted. Groups of two or three separate objects are created to represent each abstraction. One of these objects represents the data and procedures as they relate to the functional purpose of the software, i.e. parameter estimation. These are known in Java parlance as the *Models* (an unfortunate terminology for the purposes of this dissertation.) The other object or objects represent the abstraction on the computer display and allow it to be manipulated visually. The user interface objects are known as *Views* and *Controllers* if two objects are used. If a single object controls the user interface it is known as a *View-Controller*, or a *UI-Delegate*. In our case, many of the objects are paired with View-Controller objects sub classed from the `JPanel` and `JDialog` classes of the Java "swing" libraries. These View-Controllers can display an object relating to the parameter estimation process and allow it to be manipulated.

These user interface objects are shown below.

Object	Paired "View-Controller" object for user interface
`MeplanUFileParameter`	`MeplanUFileParameterPanel`
`ModeSpecificConstantDifferenceM`	`ModeSpecificConstantDifferenceJPanel`
`ModeSpecificConstantM`	`ModeSpecificConstantJPanel`
`Ratio`	`RatioPanel`
`TripCount`	`TripCountPanel`
`ODMatrix`	`ODMatrixUI`
`TripLengthDistribution`	`TripLengthDistributionUI`

9.7.7 *MEPLAN Files package*

The modelled values of the `Targets` need to be read from MEPLAN output files, and the values of the `Parameters` need to be written to the MEPLAN input files. This is accomplished through the tjmodel.meplan.files package of objects that was developed as part of this dissertation research. This package contains a `ScenarioYear` which represents a particular MEPLAN directory. The `ScenarioYear` reads in MEPLAN data files and keeps them in memory. Before the model is run any changed files in a directory are rewritten to disk by `flushing` a `ScenarioYear`. After the model is run the previously read information is discarded by `invalidateing` the `ScenarioYear`.

A MEPLAN file consists of "groups" of information (for details, see ME&P 1988). Within a `ScenarioYear` each file that is read in is divided into objects representing each `Group`. Most files can be represented by a series of generic `Groups`, but a special object called `TAM1Group` has been developed to manage the large amount of zone-pair information contained in the first group of a TAM.DAT file.

The `Parameters` and `Targets` communicate with the `Groups` through instances of the `MEPLANVariable` object. A `MEPLANVariable` is a single number in a MEPLAN file. A `Parameter` usually appears more than once in a MEPLAN file, and a `Target` is usually a function of several numbers in one or more MEPLAN files. Thus when a `Parameter` changes it will use several `MEPLANVariables` to modify a `Group`, and when a `Target` needs to recalculate its modelled value it will also use several `MEPLANVariables`.

9.7.8 Calibration Strategy User Interface

The `CalibrationStrategyUI` is the top level frame object that contains the main window of the application. It is shown in figure 22. The main window is divided into two parts. The left panel shows the parameters, while the right panel shows the targets.

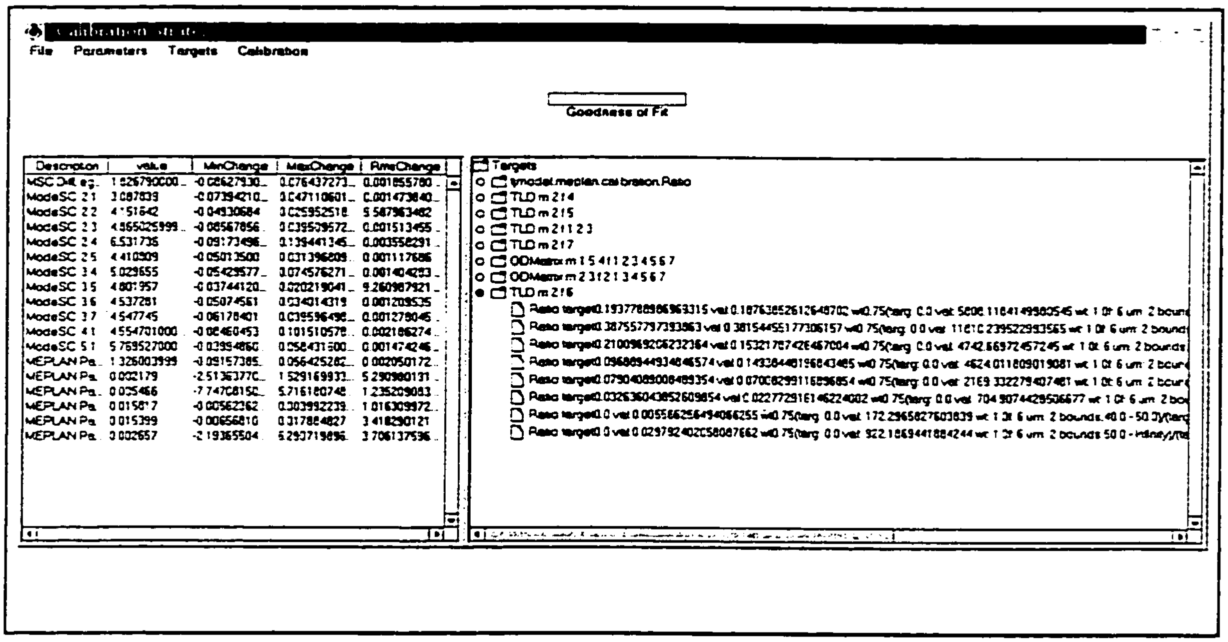


Figure 22: Calibration application

9.7.8.1 Parameter panel

The left side of the application window shows the parameters that are to be modified in the search process. The first column briefly describes the parameter. The second column shows each parameters' current value.

The third, fourth and fifth columns show the magnitude of the rows of the sensitivity matrix X_c. The third and fourth column show the minimum and maximum values in the row, being the most that this parameter would change (decrease for column 3, increase for column 4) if the weight of a single target was changed by 100% (assuming the linearization implied by equation 24 remains accurate.) The fifth column shows the root-mean-square average of the values in that row, providing a measure of the typical size of the entries in that row.

9.7.8.2 Target panel

The right side of the application window shows the targets that the search procedure is attempting to match. These targets are shown in a tree view. Targets that are not part of a `DialogableTargetGroup` are placed in the tree view according to their software Class, so all of the mode split targets, being `Ratios`, are together. Targets that are part of a `DialogableTargetGroup` show up together. In the figure the tree branch for the flow 6 trip length distribution has been expanded, showing the individual targets that make up the histogram of the `TripLengthDistribution`.

9.7.8.3 File Menu

Under the File menu, it is possible to select the Scenario. This sets which MEPLAN directory to work in and also sets up the name and location of the semaphore files that will be used to control the model running process on the separate computer (or separate process.)

There are also options to Save and Load the strategy. These selections store most of the `CalibrationStrategy` object into a file. All of the `Targets` and `Parameters` are stored as part of the `CalibrationStrategy` as well as the current modelled values for the targets values and the **F** matrix. Thus the calibration process can be stopped, saved, and resumed at a later time, an important consideration since estimating the 18 parameters in the Sacramento MEPLAN model takes weeks of computer time.

9.7.8.4 Parameters menu

The parameters menu contains menu items for creating a new instance of each target type (`ModeSpecificConstantM` "New Mode Constant", `ModeSpecificConstantDifferenceM` "New Mode Const Diff" or `MEPLANUFileParameter` "New Generic Parameter".) It also contains a menu item to "Get Current Values", which goes into the currently selected MEPLAN `ScenarioYear` and changes the values of the `Parameters` in the estimation application software to match the values in the MEPLAN input files. A menu item called "Edit Selected" will create and display the appropriate View-Controller object for the currently selected parameter.

Two menu items are used to inspect the $\boldsymbol{X}_c$ matrix. The first, "View Target Influences" creates or re-creates columns 3, 4 and 5 in the description of the parameters, to show the sensitivity of the parameter values to changes in the weights. The second, "Most Influential Targets", brings up a new dialogue box showing the 15 targets associated with the 15 weights that the selected parameter is most sensitive to, and the associated entries in the row of $\boldsymbol{X}_c$. Essentially this dialogue shows the 15 elements of the row of $\boldsymbol{X}_c$ that have the largest absolute value. These 15 targets can be compared to the third, fourth and fifth column of the parameter table to see which targets are "pulling" the strongest on the parameter value in either direction.

9.7.8.5 Target panel

The target menu contains menu items for creating new targets of two types - a `TripCount` or a `Ratio` target containing two `TripCounts`. Two menu items are for creating new groups of targets, in the form of an Origin Destination (`ODMatrix`) matrix describing the travel patterns for a particular aggregation of modes and flows, or in the form of a `TripLengthDistribution` showing the length of travel patterns for a particular aggregation of modes and flows. There is a further menu item to edit a target or a group of targets.

Two menu items are used to inspect the X_c matrix. The first, "Sensitivity of Coeff.", shows a column of X_c corresponding to a selected target. This shows the relative changes in the different coefficients if the weight of the selected target was changed.

The second, "Sens. of Other Targets", computes using equation 34 the amount the goodness of fit in other targets will change in response to a change in the weight of the selected targets. The weighted values are then sorted, so that the 15 with the largest (absolute value) change in weighted goodness of fit can be shown. Hence the modeller can see where goodness-of-fit will be sacrificed in order to achieve a better goodness-of-fit in the selected target, and where goodness of fit will be improved along with the improvement in the selected target. This menu item would have been used to identify the lack-of-fit problems described in section 9.2.1 and 9.2.2. Hence this single menu item could have been used to identify the model design issues that contributed to lack-of-fit and point to the design changes that substantially improved the model.

9.7.8.6 Calibration menu

The Calibration menu contains 4 items. The first, "start calibrating", starts a new thread to search for optimum parameter values. The use of multitasking in this regard makes it possible to still use the rest of the application while the search procedure is progressing. This is important, as with 18 parameters and a 200MHz Pentium computer it takes approximately 8 hours to do the linearization of the F matrix for the Sacramento model, and weeks to converge.

The second menu item, "stop calibrating" will send a message to the calibrating thread that it should stop at the end of the current model run.

The third and fourth menu items inspect the information matrix. The "View Covariance Matrix" shows the information matrix on the parameters (equation 27). The "View Correlations" matrix transforms the information matrix so that the diagonals are 1.0 and the off-diagonals show the correlations between the parameters.

9.8 Discussion

The calibration software was able to find a model that better matched the targets than the hand calibration. This is not surprising, as it was built with all the knowledge gained from the systematic hand calibration, and could combine that intuition with the tireless determination of a software algorithm.

How well the exact same algorithm will work for other models, even other MEPLAN models, remains to be seen. Certainly future modellers will want to use the software produced here, but it will also require modification to respond to the unique needs of each modelling effort. The "Object-Oriented" programming paradigm used here should allow the software to be easily extended to match the needs of future projects.

The search algorithm itself is a well respected local search algorithm modified to deal with the problem of repeatability in numerical modelling results. It seems that this algorithm should be broadly applicable to most situations where a local search is required.

The search algorithm is only locally convergent. There are currently a number of reasonable methods for searching for global optimums. These were not pursued for this project since MEPLAN generally requires a bit of manual "tweaking" to get it to run with vastly different parameter values. A local search process makes smaller movements and it was possible to monitor MEPLAN to determine when to adjust MEPLAN's process for solving the reduced form of the model.

It is also important to note that the global search algorithms rarely converge to a unique solution with Order 2 convergence. Generally, a global search procedure concludes with a version of Newton's method just to ensure a final set of parameters free from "random" effects. Thus the current search algorithm could enjoy widespread use even if a global search algorithm is also added to the software for future (non-MEPLAN) models.

9.8.1 Use of estimation software to investigate interdependencies

During the course of the automated search, the algorithm seemed to get stuck at a certain set of parameters. Convergence is difficult to ascertain given the numerical repeatability problems in the MEPLAN software, but it seemed that the parameter values were fairly stable and the goodness of fit was not improving. Nonetheless, there was a significant lack-of-fit in two of the trip length distributions. The $\mathbf{X}_c$ matrix was investigated using all of the menu items described above. In particular, the various targets in the suspect trip length distributions were selected and the sensitivity of the coefficients and other targets to the weights of these targets were investigated. The correlations were also investigated. The results made little sense — the relationships between some parameters and these targets seemed way too high given what was known about the structure of the model. The sensitivity of other targets also made little sense - the other targets in the trip length distribution histogram should have been sensitive to the weight of one histogram target, yet these sensitivities were smaller than many of the mode choice targets and were not always in the right direction. This inspection of the $\mathbf{X}_c$ matrix showed that there was something wrong with the search process — in particular the relationship between the spatial allocation parameter and the trip length distribution was not being uncovered. Once this was diagnosed, it was only necessary to increase the perturbation size of a location choice dispersion parameter so that it was large enough to overcome the numerical inaccuracies in the MEPLAN software and discover the larger underlying relationship and expose it in the numerical calculation of the derivatives. The calibration process was interrupted, the perturbation size increased, and the process could then converge on the appropriate parameters values.

This shows that the numerical information obtained in the search process and the ability to inspect that information in practical ways can guide any necessary manual interventions.

Figure 21 shows that, at convergence, there is some lack-of-fit in the home-to-school trip length distribution. Figure 23 shows an investigation of this using the "Sens. of other Targets" menu items. The top window shows the other targets that would be most affected by a change of the weight of the shortest trip band in figure 21, the bottom window shows that the target that has been selected (and is being investigated) is the 0 min-5 min bin of the trip length distribution.

Table 11 describes some of the interdependencies that are shown cryptically in the top window of figure 23, and offers some possible explanations for these interdependencies. The first column shows the other target that would be affected by a change in the weight of the selected target. The second column shows whether the match of the modelled value to the observed value would get better ('B') or worse ('W') when the weight of the selected target's was increased. The third column offers a possible explanation as to why this target might be interelated with the selected target.

There are a wide variety of different targets shown in table 11. This show the richness of the model — in that the search for a good fit in the home-school trip length distribution affects and is affected by other trip length distributions, mode split targets and even the details of the origin destination matrix.

The intra-zonal trip making behaviour in Davis already caused a redesign of the model to make the location choices of faculty and students exogenous during the manual calibration. The $\mathbf{X}_c$ matrix indicates, in the last row of table 11, that the trip behaviour in Davis is still strongly affecting the overall calibration. A specific mode choice submodel of Davis student household might improve the overall model substantially. Thus, *the Weight Sensitivity Matrix from the calibration procedure can help to guide changes in the model design.*

The "W" entries in Table 11 show the targets that "work against" the improvement in the selected target. One can imagine the targets connected as levers, and increasing the weight of one target is analogous to "pushing down" on that target. The "W" entries show which other targets will "pop up" when one "pushes down" on the selected target. We see that the other targets that are working against an improvement in the goodness of fit of short home-school trips are the mode choice behaviour of low income households, the number of home-school trips in other bands, and the number of short trips of different types. Thus we might consider, for example, investigating the physical locations where low income households are predisposed to make short educational trips and determine whether the representations of land use patterns and alternative modes is accurate in these places. (Inaccurate representations of alternative modes in certain zones might be the "fulcrum" that causes the mode choice behaviour of low income households to "pop up" when one "pushes down" on the trip length distribution for home-school trips.) Hence again, we see that *the Weight Sensitivity Matrix from the calibration procedure can help to guide changes in the model design.*

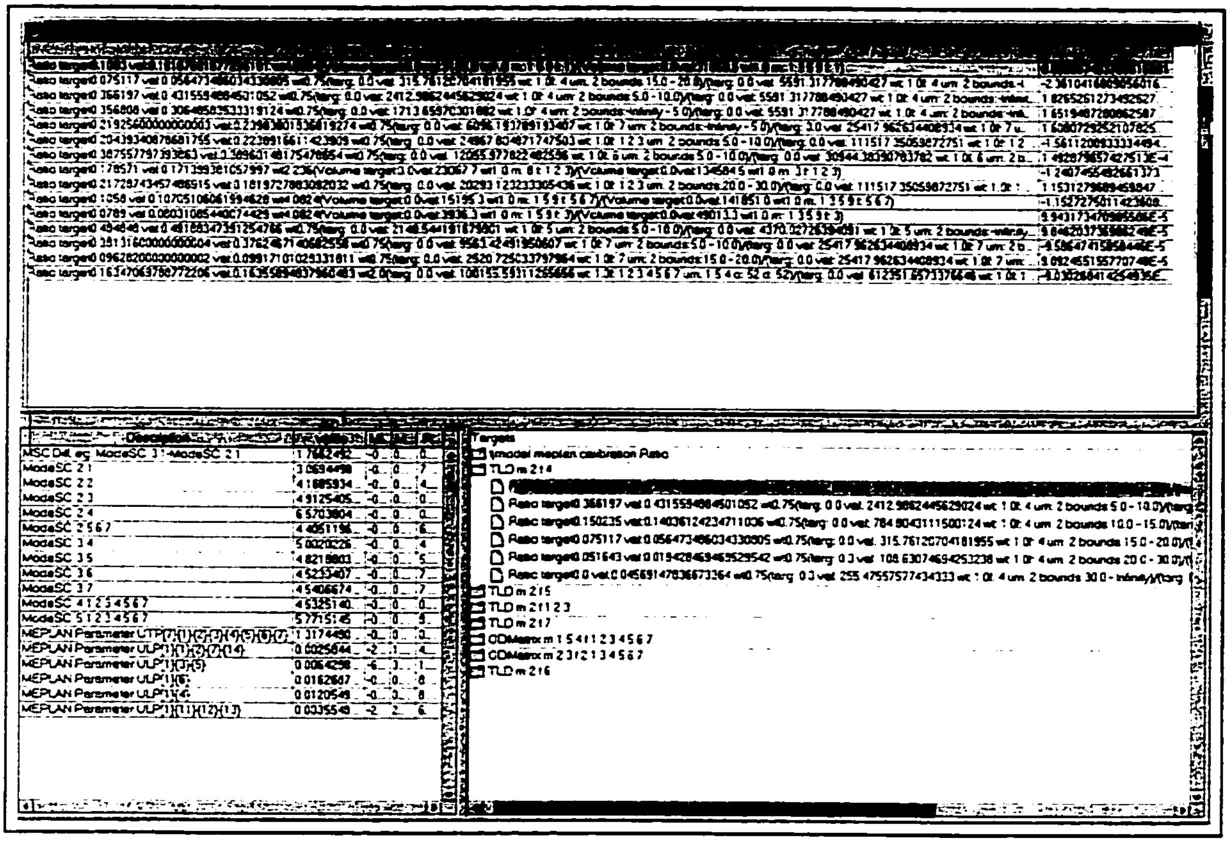


Figure 23: Investigating lack-of-fit using the $\mathbf{X}_c$ matrix

Target	B/W	Possible Explanation
Low income mode split to non-auto for work trips	W	Weight on this target is high, so this could just be an artefact of the repeatability problems in the model. Otherwise, this would be a complex interaction through the location choice of low income households, choosing locations close to schools but with good transit access
TLD for home school trips, band 15 min- 20 min	W	If more trips are pulled into the shortest band through the increase in weight, then some of those trips will be pulled out of this longer band
TLD for home school trips, band 5 min - 10 min	W	Pulling more trips into the shortest band to improve the goodness of fit will also pull more trips into the second shortest band, worsening the goodness of fit.
TLD for home school trips, band 0 min - 5 min	B	This is the target being investigated. Certainly increasing the weight of this target will make the goodness of fit better for this target
TLD for other-other trips, band 0 min - 5 min	W	This is probably a complex interaction between the various Factors competing for space. Moving activities around to

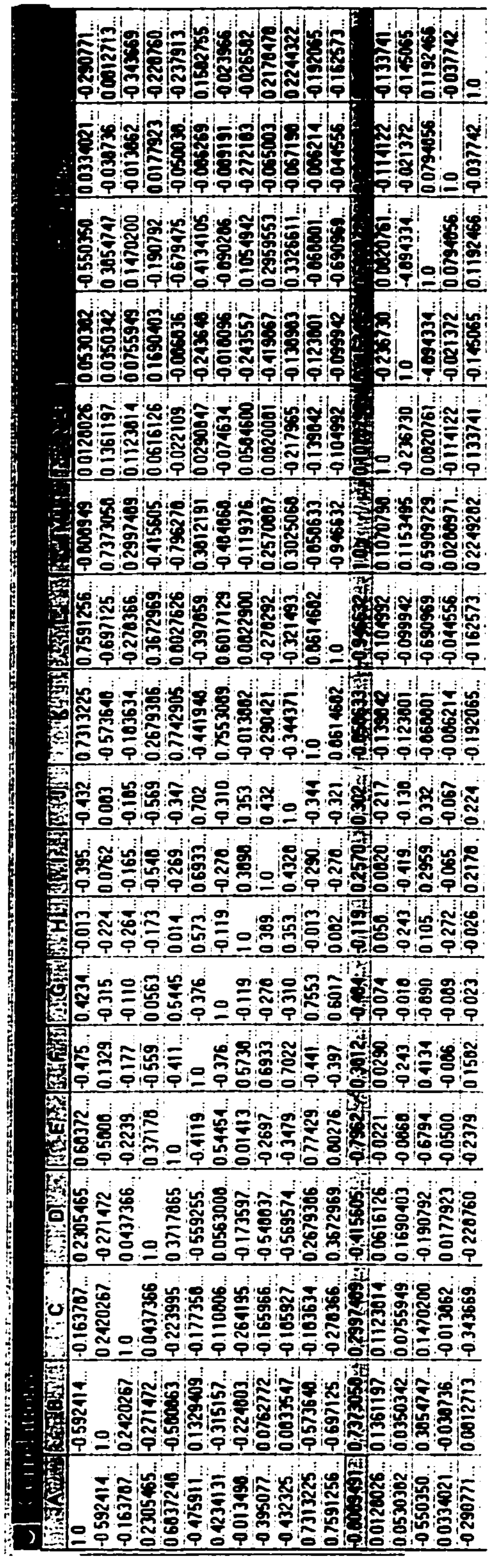
		increase the number of very short home-school trips will also increase the number of very short other-other trips.
TLD for home-work trips, band 5 min - 10 min	B	Moving households around to increase the number of short home-school trips also affects the distance of home-work trips.
TLD for home-other trips, band 5 min - 10 min	W	Moving households around to increase the number of short home-school trips also affects the distance of home-other trips.
Number of trips by non-auto modes within the city of Davis	B	Many educational trips are made in Davis. Moving households around to make these trips shorter can also lead to changes in mode choice.

Table 11: The effects on other targets of increasing the weight of the shortest bin on the home-school trip length distribution. "B" = better, "W" = worse.

The $\mathbf{X}_c$ contains over 117000 entries, and there are over 42 million cross relationships between targets and target weights. The software menus described allow these to be investigated by showing the largest items. A reasonable strategy is to select the targets where the lack-of-fit is large (either numerically large or large in light of the purpose and use of the model) and investigate the cross relationships with other targets and parameters. These will lead to various hypotheses about the cause of the error. These hypotheses may lead to changes in the form of the goodness of fit function, or a design change in the model. Or perhaps the model and goodness of fit function may be seen to be adequate, and the modeller then may choose to increase the weight of the selected target, already knowing approximately how this will affect the overall model when it is re-estimated with the new weight.

9.8.1 Correlation matrix

The correlation matrix is computed by the software and is shown in figure 24, with rows and columns corresponding to the parameters in table 12.



Correlations

A	B	C	D	E	F	G	H	I	J	K	L	M	N				
1.0	-0.592414...	-0.163787...	0.2305465...	0.68372...	-0.475...	0.4234...	-0.013...	-0.395...	-0.432...	0.7313225...	0.7591256...	-0.808949...	0.0128026...	0.0530382...	-0.550350...	0.0334021...	-0.290771...
-0.592414...	1.0	0.2420267...	-0.271472...	-0.5808...	0.1329...	-0.315...	-0.224...	0.0762...	0.083...	-0.573648...	-0.697125...	0.7373058...	0.1361197...	0.0350342...	0.3854747...	-0.038736...	0.0812713...
-0.163787...	0.2420267...	1.0	0.0437366...	-0.2239...	-0.177...	-0.110...	-0.264...	-0.165...	-0.185...	-0.183634...	-0.278366...	0.2997489...	0.1123814...	0.0755949...	0.1470200...	-0.013862...	-0.343669...
0.2305465...	-0.271472...	0.0437366...	1.0	0.37178...	-0.559...	0.0563...	-0.173...	-0.548...	-0.569...	0.2679386...	0.3672969...	-0.415605...	0.0616126...	0.1690403...	-0.190792...	0.0177923...	-0.228760...
0.6837248...	-0.588863...	-0.223995...	0.3717865...	1.0	-0.411...	0.5445...	0.014...	-0.269...	-0.347...	0.7742905...	0.8027626...	-0.796278...	-0.022109...	-0.086836...	-0.679475...	-0.050038...	-0.237913...
-0.475911...	0.1329409...	-0.177358...	-0.559255...	-0.4119...	1.0	-0.376...	0.573...	0.6933...	0.702...	-0.441948...	-0.397859...	0.3812191...	0.0290847...	-0.243648...	0.4134105...	-0.086269...	0.1582755...
0.4234131...	-0.315157...	-0.110806...	0.0563008...	0.54454...	-0.376...	1.0	-0.119...	-0.278...	-0.310...	0.7553089...	0.6017129...	-0.484868...	-0.074634...	-0.018096...	-0.890286...	-0.089191...	-0.023966...
-0.013498...	-0.224803...	-0.264195...	-0.173597...	0.01413...	0.5738...	-0.119...	1.0	0.3898...	0.353...	-0.013882...	0.0822900...	-0.119376...	0.0584600...	-0.243557...	0.1054942...	-0.272183...	-0.026582...
-0.395077...	0.0762772...	-0.165966...	-0.548837...	-0.2697...	0.6933...	-0.278...	0.389...	1.0	0.432...	-0.290421...	-0.278292...	0.2570887...	0.0820081...	-0.419867...	0.2959553...	-0.065003...	0.2178478...
-0.432325...	0.0833547...	-0.185927...	-0.569574...	-0.3479...	0.7022...	-0.310...	0.353...	0.4328...	1.0	-0.344371...	-0.321493...	0.3025068...	-0.217965...	-0.138983...	0.3326611...	-0.067198...	0.2244322...
0.7313225...	-0.573648...	-0.183634...	0.2679386...	0.77429...	-0.441...	0.7553...	-0.013...	-0.290...	-0.344...	1.0	0.8614682...	-0.850633...	-0.139842...	-0.123801...	-0.868801...	-0.086214...	-0.192065...
0.7591256...	-0.697125...	-0.278366...	0.3672969...	0.80276...	-0.397...	0.6017...	0.082...	-0.278...	-0.321...	0.8614682...	1.0	-0.946632...	-0.104992...	-0.099942...	-0.690969...	-0.044556...	-0.162573...
-0.8089491...	0.7373058...	0.2997489...	-0.415605...	-0.7962...	0.3812...	-0.484...	-0.119...	0.2570...	0.302...	-0.850633...	-0.946632...	1.0	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
0.0128026...	0.1361197...	0.1123814...	0.0616126...	-0.0221...	0.0290...	-0.074...	0.058...	0.0820...	-0.217...	-0.139842...	-0.104992...	0.1070790...	1.0	-0.236730...	0.0820761...	-0.114122...	-0.133741...
0.0530382...	0.0350342...	0.0755949...	0.1690403...	-0.0868...	-0.243...	-0.018...	-0.243...	-0.419...	-0.138...	-0.123801...	-0.099942...	0.1153495...	-0.236730...	1.0	-4.894334...	-0.021372...	-0.145065...
-0.550350...	0.3854747...	0.1470200...	-0.190792...	-0.6794...	0.4134...	-0.890...	0.105...	0.2959...	0.332...	-0.868801...	-0.690969...	0.5909729...	0.0820761...	-4.894334...	1.0	0.0794856...	0.1192466...
0.0334021...	-0.038736...	-0.013862...	0.0177923...	-0.0500...	-0.086...	-0.089...	-0.272...	-0.065...	-0.067...	-0.086214...	-0.044556...	0.0288971...	-0.114122...	-0.021372...	0.0794856...	1.0	-0.037742...
-0.290771...	0.0812713...	-0.343669...	-0.228760...	-0.2379...	0.1582...	-0.023...	-0.026...	0.2178...	0.224...	-0.192065...	-0.162573...	0.2249282...	-0.133741...	-0.145065...	0.1192466...	-0.037742...	1.0

Figure 24: Correlation matrix

Letter	Parameter
A	Difference between MSC for HOV and MSC for SOV, for work trips
B	MSC for SOV for low income work trips
C	MSC for SOV for mid income work trips
D	MSC for SOV for high income work trips
E	MSC for SOV for home-school trips
F	MSC for SOV for all other trips
G	MSC for HOV for home-school trips
H	MSC for HOV for home-shop trips
I	MSC for HOV for home-other trips
J	MSC for HOV for other-other trips
K	MSC for transit
L	MSC for bicycle
M	Mode choice dispersion parameter
N	Location choice dispersion parameter for OFSRV-IND, GOVT and MANUF
O	Location choice dispersion parameter for HEALTH and OFSRV-RES
P	Location choice dispersioin parameter for EDUCATION
Q	Location choice dispersion parameter for RETAIL
R	Location choice dispersion parameter for households (HH-LOW, HH-MID and HH-HIGH)

Table 12: Parameters adjusted by the software

There are fairly large number of numbers greater (in absolute value) than 0.6 in the matrix. This is empirical evidence confirming the experience that "hand calibration" is notoriously difficult. During hand calibration the theory of the model suggests some strong relationships between certain parameters and certain targets. If there is mismatch in a target the theory suggests which parameter can be modified to achieve a better fit in that target. The correlation matrix shows that after a target has been matched by the adjustment of one parameter many other targets will have also been affected, and these targets also need to be considered. The approach in the hand calibration (and in previous hand calibrations, Hunt 1994) has been to identify which targets are directly affected by each parameter in turn, and adjust that parameter according to the goodness of fit in those targets. But the very nature and purpose of a land-use/transport model (which is to consider a wide range of interactions and secondary effects) means that each parameter will influence all the model outputs. The correlation matrix shows that these interdependencies do not cancel each other, hence when all parameters have been considered individually the secondary effects and interactions make it necessary to revisit the earlier calibrations in a circular procedure.

The automated procedure can consider the complex interdependencies between all of the 18 parameters and over 6000 targets and use a numerical procedure to adjust all the parameters at once to overcome these interdependencies.

Part 4.

Conclusions

10 Conclusions

10.1 Parameter estimation

The taxonomy of methods for dealing with the complexity of the modelling system in estimating parameters has proven to be useful. The investigation of statistical theory showed the inadequacy of "piece wise" estimation, and hence the importance of performing an overall estimation.

Performing the overall estimation using the actual modelling software had advantages and disadvantages. With the estimation runs essentially identical to the policy analysis runs, the overall estimation process served to validate the software implementation of the model and give confidence that the model will reproduce the key outputs to be viewed by policy makers. But the modelling software has complex iterative procedures to solve for the output variables (i.e. to solve the "reduced form"), and these procedures do not solve for the output variables with suitable accuracy to enable a complete search for the best parameters. Further, it takes a long time (about 30 minutes on a Pentium 200MHz computer with 32 MB of RAM) to solve for the output variables, causing the overall parameter search to take weeks. These weeks are much shorter than the months of difficult skilled labour required to manually estimate the parameters, but nonetheless these weeks may be unnecessary. A more theoretically correct quasi-likelihood function could be written in terms of the reduced form of the model. This would eliminate the need for complex numerical procedures for determining the goodness-of-fit, considerably shortening the time needed to estimate the parameters. A worthwhile future research project would be to perform an overall estimation of a MEPLAN model using a goodness-of-fit measure based on the reduced form of the model. The standard MEPLAN software could not be used for this, however.

Land use/transport models have been criticized for using "synthetic" data — where the inputs for one submodel are calculated by another submodel. The statistical theory shows that this criticism is valid if the parameters are estimated by considering submodels separately. But the criticism can be addressed by performing an overall estimation to eliminate the biases that may be introduced. Further, the nested logit example in chapter 8 shows that "synthetic data" occurs even in simple models. Any intermediate variable that might represent something behavioural or physical could be considered "synthetic". Certainly it is desirable to have data representing every quantity in a model, but that is seldom realistic unless the behavioural nature of modelling is sacrificed. (Imagine rejecting all random utility models because we don't normally have any direct measure of "utility"!) The appropriate way to deal with intermediate variables for which no measured data exists is to ensure that the variable really is "intermediate" by considering several sub-models at once. The appropriate way to deal with intermediate variables for which measured data does exist is to include that data in the goodness-of-fit function for the overall estimation.

The goodness-of-fit functions used at different stages in the calibration should conform to statistical theories and hypothesis regarding the sources and sizes of errors. But the "best" model is the one that is most useful for policy analysis, so the purpose of the model must also be taken into account. A reasonable procedure is to base the initial goodness of fit function on statistical theories and error hypothesis, but then adjust the goodness-of-fit function manually based on the analyst's understanding of what errors are important given the way the model will be used. In a least-squares goodness-of-fit function, these adjustments can be made by changing the weights associated with each error measurement.

Bayesian theory can help to justify why in sequential estimation some parameters are estimated in lower level sub-models, then "fixed" in the overall estimation, while other parameters are estimated or adjusted in the overall estimation. Each relationship in the model together with the observed data can provide information on the likely value of certain parameters. Some of these relationships and data can provide a strong indication of the value of a parameter, while others might only provide a weak indication of the value of the parameter. In Bayesian terms, the "posterior distribution" from an estimation contains all the information on a parameter value from that estimation. If the posterior distribution from a submodel estimation has a smaller variance than the posterior distribution from the overall estimation, then the submodel estimation informs the parameter more than the overall estimation and hence that parameter could be fixed in the overall estimation. Often it is obvious from theory what primary relationships in the data will most inform a parameter value.

Simultaneous estimation can consider all of these relationships together and determine the "best" value for the parameter given all the available data and relationships. But full simultaneous estimation is unrealistic for large, complex urban systems models. Bayesian Sequential Estimation may prove to be useful in future exercises. In Bayesian Sequential Estimation, the posterior distribution of the parameters is calculated for each submodel. This posterior distribution can be used as a "prior distribution" for the higher levels of estimation, informing the values of the parameters while still allowing the data at the higher levels to inform the values as well. This procedure holds considerable promise for parameters shared between submodels, although it would still be useful to fix many parameters during the highest level estimation just to manage the complexity of the problem.

The Weight Sensitivity Matrix developed in this research shows how parameters will change if a weight on one target is changed by a small amount. It can also be used with the linearized model to show how the goodness-of-fit of all targets will be affected by a change in a single weight. The information in this matrix and the methods of inspecting it (as embodied in the software) may prove to be an especially valuable tool for calibrating urban systems models. The ability to easily identify conflicting targets would have been especially useful during the parameter search for the Sacramento model; as two design changes were made specifically in response to lack of fit on combined targets. Hunt (1994) states for the Naples MEPLAN model:

> Another problem that arises because of the extensive scope of the model relates to making judgements about the consistency of the observed data ... given the perspective of the model formulation. ... Such inconsistency is likely to be in part a result of error in the observations, and in part a result of problems within the framework.

The Weight Sensitivity Matrix and the software tools for inspecting the matrix provide easy methods for identifying such inconsistencies and hypothesizing as to their cause. The automation of the estimation process means it is no longer so "complex, slow, and expensive" that it "curtail[s] consideration of alternative forms for the model components" (Hunt, 1994). This should allow for better models, shorter development times and lower costs.

It would be difficult to show for certain whether the semi-automated method developed here would lead to better models than a more manual method. Multiple teams of modellers could be selected from across the world, and some could be told to use the automated method while others could be told to rely on other methods. The teams could be given similar data and budgets, and afterwards their resulting models could be compared. This is similar to the strategy employed by the University of California at Davis for comparing modelling frameworks. The researchers at the University of

Calgary were asked to develop a model of the Sacramento region based on the MEPLAN framework, while other researchers were asked to construct models based on other frameworks. The results of these are now being compared, and it seems that some conclusions can be made regarding the strengths of the different frameworks. The same approach was taken during the ISGLUTI (International Study Group on Land Use Transport Interaction) study (Webster *et al*, 1988), which is highly cited as an informative comparison of modelling frameworks. It would be useful to have a similar study that compares calibration mechanisms.

Finally, it should be noted that the parameters obtained through the automated search were unique given the objective function, while the parameters obtained through manual searches are highly dependent on the decisions, persistence and skill of the model builder (as was demonstrated in the early manual calibration for the Sacramento model, before it was discovered that the mode choice sensitivity parameter could be changed in linear combination with the other mode choice parameters to improve the goodness of fit.) As well, the goodness of fit from the automated procedure was substantially better than the goodness of fit achieved in the manual search.

There seems to be a dichotomy of approaches for building accurate models. One approach, which might be called "statistical", involves creating a mathematical model with only a few equations, and estimating the parameters of that model based on a statistical likelihood function. This method leads to simple models that can provide great insight regarding the details of part of the system, but that can rarely be used for practical forecasting because only simple models can receive this kind of in-depth statistical treatment. Papers that describe these models are dominated by mathematical equations, likelihood functions and statistics. The other method, which might be called "demonstration by simulation", involves constructing complex computer simulations based on rules-of-thumb and plausible but untested relationships. These simulations are then run and may be tweaked and adjusted if their results do not seem reasonable. Occasionally, they are "validated", but this usually involves no formal statistics — only an informal comparison of results against expectations. Papers that describe these models are dominated by figures showing the complexity and beauty of the model and descriptions of the computer software that runs the simulation.

There is a surprising lack of work in between these extremes — that is, a complex computer simulation that has been made accurate by estimating parameters through statistical means. This dissertation helps to fill a void that probably shouldn't exist.

10.2 Land Use Models in Transportation Planning

There a wide variety of different urban modelling frameworks, each with a different emphasis and with different strengths. The MEPLAN framework is flexible, and models that use the framework can be designed to represent many of the different urban economic and transportation issues that policy-makers may want to consider. Its use of behavioural logit models is theoretically and practically appealing.

Nonetheless, there are certain things that can not be represented within a MEPLAN framework. MEPLAN emphasizes market economics and market equilibrium. Dynamic processes can be difficult to represent properly, especially if a process involves some aspect of "foresight" (where actors make decisions based on what they know about the future.) Market failure (due to incomplete information or excessive government regulation) cannot always be represented.

The MEPLAN framework represents categories of activity. The diversity of individual actors within a category is only represented through single error terms on utility functions. This cannot represent how actors with similar preferences and needs tend to group together and adjust their environment to serve their needs. Urban systems are known for how different neighbourhoods adopt various characteristics based on race, income, ethnicity or occupation. Transportation services and urban spatial structure evolves to serve the unique characteristics that develop in each neighbourhood. It's difficult to represent this sort of clustering within an aggregate model (like a MEPLAN model), unless the reasons for the clustering are identified in advance. The well-respected urban commentator and observer, Jane Jacobs, in "The Economy of Cities" (1969), argues that the entire strength of urban economies rests on the great diversity of individual businesses, firms and people and how these actors search amongst each other to find (and locate beside) the best suppliers given their own, unique, needs. A model that does not represent this might be missing the *raison d'être* of cities.

Further work should be done to develop a microsimulation based model of an urban economy that can represent this fundamental clustering between complementary activities without directly representing all the differences between actors. The unique characteristics of actors could be represented by sampling a variety of random variables from distributions, some of which could represent measured attributes (such as parameters of consumption functions) while others would represent the remaining attributes and characteristics that are not of interest from a policy perspective. The actors could arrange themselves spatially to match with complementary activities.

This type of model would show the fundamental strength of microsimulation approaches — the ability to represent local, micro-level diversity.

A possible way forward is to develop a software system that embodies the aggregate "factor" representation of an urban economy in the same way the MEPLAN does, but to design the software so that disaggregate, microsimulation submodels can replace the aggregate submodels for different factors as they are developed.

Another possible way forward is to develop the simplest possible representation of a diversity of actors co-locating spatially to match with complementary actors. Once such a model is developed, the actors can be divided into categories that match the "factors" in a MEPLAN model, and the range of attributes of the different factors can be redefined to represent measured variables and policy inputs.

10.3 Model Development using the MEPLAN framework for Sacramento

This project has demonstrated that the MEPLAN framework can be used to develop a meaningful and useful model of urban land use and transport interaction in a United States situation.

This is perhaps not too surprising given the flexibility of the framework and its history of successful application around the world in various contexts, but it is nevertheless a confirmation. It should foster a greater appreciation and recognition of the possibilities regarding land use and transport modelling in the United States, which would seem to be particularly relevant given the recent attention and legislation concerning land use and transport interaction.

The specification of policy inputs is more direct than with standard transport models: the effects of a potential land release policy are investigated by specifying land to be released as a direct input to the model, whereas with a standard transport model the impacts of the land release policy on distributions of employment and population would have to be anticipated and the results of this anticipation input to the model.

The model was designed around the available data, and unanticipated trends and inconsistencies in the data led to changes in the model design. Some of these changes were improvements, others might be considered compromises. This is generally the case with MEPLAN - the flexibility of the framework allows models to be designed to address the policies of interest while respecting the practical limitations of available data.

The flexibility means that there are often several ways to approach a given problem regarding data or lack of model fit, and experience and insight with urban systems and modelling in general and with MEPLAN specifically are particularly valuable. This project has formalized the parameter estimation process to some degree, allowing such insight and experience to be applied in the selection of the model structure, where it is most valuable. Further, the Weight Sensitivity Matrix from the formal overall estimation procedure has proved useful for suggesting any changes in the structure of the model.

The lack of complete and consistent information on land prices and land quantities was not ideal - but it was not a 'pathological' fault making it impossible to develop a useful model. Land prices in a land market model are analogous to travel times in transportation demand models; and transportation demand models have been calibrated and applied for years without complete information on all travel times for all links. A model and calibration procedure should be designed to work around incomplete or inconsistent data to identify the underlying relationships and trends.

As with all models, there is room for further improvement in the Sacramento model:

- the Stone-Geary or Klein-Rubin form of direct utility consumption functions for households could be implemented, allowing the model to predict how consumption by households will change as prices change;
- the incremental allocation models that represent the urbanization and development of land should be more rigorously calibrated to observed time series data; and
- more complete data on floorspace (building stock) and on the current distribution of land uses could be used to more accurately represent the market for space.
- disaggregate data on firm location decisions could be investigated to further understand the way industrial activity arranges itself to reduce transportation costs; and the model's design or parameters could be changed as a result.
- the spatial allocation process could be changed to use the travel characteristics from all of the simulated travel periods, instead of just the AM 1hr peak period. This could allow spatial patterns to evolve in response to different travel conditions at different times of the day.
- goods movement could be added to the model, so that the impact of shipping costs on firm location choice can be represented directly.
- the size and structure of the overall economy could be allowed to vary in response to modelled variables, instead of being constant across all scenarios.

10.4 Scenario evaluation

Cross-movement of different industry types is prominent in the scenario predictions. Activities locate so as to be well positioned with regard to inputs and consumers, and the differences in relative consumption by different industries and households leads to closely linked activities 'following each other' around the urban area. This suggests that more emphasis should be placed on firm location, through further research, data collection and enhancements in the model.

The resulting model of Sacramento is a potentially useful tool for policy analysis because it allows policy makers to investigate in a consistent manner how both transport and land use policies influence both land use and transport conditions, including land values, the spatial distributions of activities and transport conditions for specific modes and facilities. The scenario results show how activity location responds to transportation policy and how transportation behaviour changes when certain land development patterns are encouraged.

The model was run with a single point forecast for each policy scenario. The future is full of uncertainty, and a single point forecast is probably of less value than an indication of the range of possible futures that might result from a policy direction and the probability that different future events may occur given a certain policy direction. A run of a computer simulation model could be considered to be a single experiment (Naylor, 1971), and in science multiple experiments are generally required before conclusions can be made. This seems especially important when one considers that urban systems are likely chaotic (as shown in Allen, 1997, and hinted at already in Christaller, 1933), and that the MEPLAN model of Sacramento itself demonstrated the "extreme sensitivity to initial conditions" associated with chaos (section 6.2.7). The input values and parameters could be perturbed according to assumed distributions, and the model run multiple times, to simulate a range of possible outcomes. The resulting experience and knowledge relating to policy evaluation and planning in the face of uncertainty would prove valuable in other models that more explicitly represent the chaotic dynamics of urban systems.

The policy evaluations were performed using the "hand calibrated" model. There were two reasons for this. First, the University of California at Davis sponsored the initial model development and expected certain scenario results before the work on automatic calibration was complete. Second, it was thought that a greater understanding of how models are used would lead to insight that could influence the automatic calibration procedure. To some extent this did happen: as scenario results were compared to each other the relative importance of different targets became apparent leading to changes in weights and changes in design. On the other hand, performing the scenario evaluations required a lot of effort, at least in comparison to their influence on the automated calibration method. It might have been better to focus on the automated calibration first, then afterwards perform only a few scenario evaluations to demonstrate the functionality of the completed model.

10.5 Project organization

The co-operation with the University of California Davis in providing the data for the modelling exercise was useful. UC Davis is currently using the model for various policy analysis, exposing the United States to the full capabilities of such models.

10.6 Software implementation

The use of Java for object oriented programming, instead of C++, made development easier but increased the run time of the estimation software. This was a desirable trade-off since the bulk of the computing time required was for the MEPLAN modelling software, not the estimation software. Java software is also platform independent, allowing the estimation software to be reused with other models on other operating systems.

The user interface for the estimation software was constructed from the Java "Swing" libraries. These libraries provide a useful tool kit of user interface components. Even so, it was surprising how much of the details of user interface functionality had to be manually created. The Visual Age for Java development environment from IBM allowed much of this to be done using "drag and drop" on screen design instead of by writing code, but it still seemed like much too much time was being spent writing software to display items on the computer screen and allow them to be manipulated. There are more advanced user interface libraries available — some of these may have been worth pursuing. Gu and Liu have recently (1999) presented a review of visualization software for urban models — a useful guide for future researchers.

The connection between the model and the estimation software through the file system (semaphores for process control, and data files for data sharing) had advantages. This allowed the modelling to occur without any modifications to the MEPLAN software. It also allowed the model to run on a separate machine and operating system than the estimation software. The resulting set of objects create an abstraction of the modelling process that should prove useful if the estimation software is deployed for other models. Nevertheless, considerable effort was required to develop the software objects that manipulate the MEPLAN data files and the semaphore files, and the file system connection did not facilitate the modern two-way "event based" connections where processes send requests to each other asking to be notified when various things happen. There are more modern standards for inter process communication that could be adopted in future work, although they would need to be supported by the modelling software as well as the estimation software.

References

Abraham, J.E. and Hunt, J.D., 1997, Specification and Estimation of a Nested Logit Model of Home, Workplaces and Commuter Mode Choices by Multiple Worker Households. *Transportation Research Record*, V1606 pp17-24

Abraham, J.E. and Hunt, J.D., 1998, Calibrating the MEPLAN model of Sacramento. *Transportation Research Board Annual Conference, Washington DC, January 1998*, paper 980649 (available on the Pre-print CD-ROM)

Abraham, J.E. and Hunt, J.D., 1999a, Policy Analysis Using the Sacramento MEPLAN Land Use Transportation Interaction Model. *Transportation Research Record* 1685:199-208

Abraham, J.E. and Hunt, J.D., 1999b, Firm Location in the MEPLAN model of Sacramento. *Transportation Research Record* 1685:187-198

Abraham, J.E. and Hunt, J.D., 1999c, Semi-Automated Calibration of the MEPLAN Model of Sacramento, 6th Computers in Urban Planning and Urban Management Conference, September 8-11, Venice, Italy.

Allen, Peter M., *Cities and Regions as Self-Organizing Systems: Models of Complexity*, Gordon and Breach, Amsterdam, The Netherlands, 1997

Amemiya, T., 1974, "The Nonlinear Two-stage Least Squares Estimator," *Journal of Econometrics*, July 1974, pp. 105-110

Anas, A., 1987, *Modelling in Urban and Regional Economics*, Harwood Academic Publishers, CmbH, Chur, Switzerland

Anas, A., and Cho, J.R., 1986, "A Dynamic, Policy Oriented Model of the Regulated Housing Market: The Swedish Prototype," Working Paper. Northwestern University, Evanston

Arnott, R.J., 1985, "HOPSIM: A Housing Policy Simulation Model," Queens University, Kingston, Ontario

Bard, Y, 1974, *Nonlinear Parameter Estimation*, Academic Press, New York and London

Bates, D. M., and Watts, D. G., 1988, *Nonlinear Regression Analysis and its Applications*, John Wiley & Sons, New York

Batty, M., and Longley, P., 1994, *Fractal Cities : A Geometry of Form and Function*, Academic Press

Brownstone, D., Small, K. A., 1989, "Efficient Estimation of Nested Logit Models" *Journal of Business and Economic Statistics*, 7(1):67-74

Christaller, Walter, 1933, *Die zentralen Orte in Süddeutschland*. Jena: Gustav Fischer (Translated (in part), by Charlisle W. Baskin, as *Central Places in Southern Germany*. Prentice Hall 1966).

Cuthbertson, K., Hall S. G. and Taylor, M. P., 1992, *Applied Econometric Techniques*, Philip Allan, Hertferdshire (Chapter 8)

Daganzo, C., and Y. Sheffi, 1977, "On Stochastic Models of Traffic Assignment," *Transportation Science* 11:243-274

de la Barra T, 1989, *Integrated Land Use and Transport Modelling*, Cambridge University Press, Cambridge, England

DeLeeuw, F., and R.J. Struyk, 1975, *The Web of Urban Housing: Analyzing Policy with a Market Simulation Model*, The Urban Institute, Washington DC

DKS Associates, 1994, *SACMET Regional Travel Demand Model Model: Model Development and User Reference Report prepared for Sacramento Area Council of Governments*

Douglas PH, 1934, *The Theory of Wages*, Crowell, Collier and MacMillan, New York, USA

Engle, R.F., F.M. Fisher, J.R. Harris and J. Rothenberg, "An Econometric Simulation of Intra-Metropolitan Location: Housing, Business, Transportation and Local Government," *American Economic Review* 52:87-97

EPA (United States Environmental Protection Agency), 1988, *Technical Methods for Analyzing Pricing Measures to Reduce Transportation Emissions*, Publication number EPA 231-R-98-006, Washington, DC

Fisk, C., 1980, "Some Developments in Equilibrium Traffic Assignment, *Transportation Research B* 14:243-255

Forrester, J.W., 1969, *Urban Dynamics*, MIT Press, Cambridge, Mass.

Fujita, M, 1987, *Urban Economic Theory: Land Use and City Size*, Cambridge University Press, Cambridge, UK

Gleick, James, *Chaos: Making a New Science*, Penguin Books, New York, 1987

Gu, K.Q. and Jin, V.X., 1999, "Data representation and visualization for land use and transport models", *CUPUM '99: Computers in Urban Planning and Urban Management on the Edge of the Millenium*, proceedings of the 6th international conference, Venice 1999. FrancoAngeli, Milan

Hartwick, P.G., and Hartwick, J.M., 1974, "Efficient Resource Allocation in a Multinucleated City with Intermediate Goods," *Quarterly Journal of Economics* 88:340-352

Herbert, J.D., and Stevens, B.H., 1960, "A Model for the Distribution of Residential Activity in Urban Areas," *Journal of Regional Science* 2:21-36

Hunt, J.D., 1994, Calibrating the Naples land use and transport model. *Environment and Planning* 21B:569-590.

Hunt, J.D., 1997, A Stated Preference Examination of the Location Choice Behavior of Retail Firms, *Transportation Research Record* 1606:25-32

Hunt, J.D. and Echenique, M.H., 1993, Experiences in the application of the MEPLAN framework for land use and transport interaction modeling. *Proceedings of the 4th National Conference on the Application of Transportation Planning Methods*, Daytona Beach, Florida, USA, May 1993, pp:723-754.

Hunt, J.D. and McMillan, J.D.P., 1995, The Edmonton MEPLAN Model: Calibration and Data Considerations. *Proceedings of the 4th International Conference on Computers in Urban Planning and Urban Management*, Melbourne Australia, July 11-14 1995, 2:97-110

Hunt, J.D., McMillan, J.D.P. and Abraham, J.E., 1994, Stated Preference Investigation of Influences on Attractiveness of Residential Locations, *Transportation Research Record* 1466:79-87

Hunt, J.D. and Simmonds, D.C., 1993, Theory and Application of an Integrated Land-Use and Transport Modelling Framework, *Environment and Planning* B 20:221-244

Jacobs, Jane, 1969, *The Economy of Cities*, Vintage Books (Random House), New York

Kelley, A.C. and Williamson, J.G., 1983, "A Computable General Equilibrium Model of Third World Urbanization and City Growth: Preliminary Comparative Statistics," in *Modelling Growing Economies in Equilibrium and Disequilibrium*, ed. by A.C. Kelley, W. Sanderson and J.G. Williamson, Duke University Press, Durham

Kim, T.J., 1979, "Alternative Transportation Modes in an Urban Land Use Model: A General Equilibrium Approach," *Journal of Urban Economics* 6:197-215

Knox, P, 1987, *Urban Social Geography: An Introduction*, Longman Scientific and Technical, Harlow, Essex, England

Koopmans, T.C., and M.J. Beckmann, 1957, "Assignment Problems and the Location of Economic Activities", *Econometrica* 1:53-76

Leontief W., 1941, *The Structure of the American Economy 1919-1939*. Oxford University Press, New York NY, USA.

Lowry, I.S., 1964, "A Model of Metropolis", RM-4035-RC, RAND Corporation, Santa Monica USA

Marquardt, D.W., 1963, An algorithm for least squares estimation of nonlinear parameters. *SIAM Journal* 11:431-441

McFadden, D., 1974, Conditional logit analysis of qualitative choice behavior. In *Frontiers in Econometrics*, Editor: Zarembka P. Academic Press, New York, USA, 105-142.

ME&P (Marcial Echenique & Partners Ltd.), 1994, *MEPLAN 2.40 User Reference Manual,* Cambridge, UK

ME&P (Marcial Echenique & Partners Ltd.), 1989, *MEPLAN Version 2.2 Urban User Guide,,* Cambridge, UK

Meyer, M.D. and Miller, E.J., 1984, *Urban Transportation Planning: A Decision-Oriented Approach*. McGraw-Hill, New York, USA.

Mills, E.S., 1974, "Mathematical Models for Urban Planning", in *Urban and Social Economics in Market and Planned Economies*, ed. by A. Brown *et al.*, Praeger, New York.

Modelistica, 1996, Application of the TRANUS integrated land use and transport model to the Sacramemento Metropolitan Region, July 1996

Naylor, T.H., 1971, *Computer Simulation Experiments with Models of Economic Systems*, John Wiley & Sons, New York

Popper, K. R., 1959, *The Logic of Scientific Discovery*, Basic Books, New York

Press, W. H., Flannery, B. P., Teukolsky, S. A. and Vetterling, W. T., 1988, *Numerical Recipes in C: The Art of Scientific Computing,* Cambridge University Press, Cambridge UK

Pyatt G and Thorbecke E, 1976, *Planning Techniques for a Better Future*. International Labour Office, Geneva, Switzerland.

Richardson, H.W., 1971, *Urban Economics*, Penguin, Harmondsworth

Sacramento Area Council of Governments, 1996, *1996 Metropolitan Transportation Plan Working Paper #3: Analysis of the Options*

Samuelson, P.A., 1952, "Spatial Price Equilibrium and Linear Programming," *American Economic Review* 42:283-303

Theil H, 1980, *The System-Wide Approach to Microeconomics*. University of Chicago Press, Chicago, USA.

Shuper, Viacheslav, 1997, "Applications of Relativistic Central Place Theory", 10th European Colloquium on Theoretical and Quantitative Geography in Rostock, Germany, 6 - 10 September 1997

von Thunen, J. H., 1826, *Der Isolierte Staat in Beziehung auf Landwirtshaft und Nationalekonomie*, Hamburg

Webster, F.V., P.H. Bly and N.J. Paulley, 1988, "Urban Land Use and Transportation Interaction," Avebury 1988.

Williams, H.C.W.L., 1977, On the formation of travel demand models and economic evaluation measures of user benefit. *Environment and Planning* 9A(3):285-344.

Wilson, A.G., 1967, "A Statistical Theory of Spatial Distribution Models," *Transportation Research* 1:253-269

Appendix A

In MEPLAN it is common to use an exponential expenditure function (e.g. function 11 in ULP[3]).

$$a_{m,i} = m_{m,i} + A_{m,i} \cdot e^{-B_{m,i} \cdot P_i} \tag{A1}$$

where:

P_i	is the price of factor i (CONSCOST$_i$)
$m_{m,i}$	is the minimum consumption of factor i in the production of factor m
$a_{m,i}$	is the actual consumption of factor i in the production of factor m
$A_{m,i}$ and $B_{m,i}$	are coefficients

The attractiveness of a zone is inversely related to the cost of producing the factor in the zone, which is generally a sum

$$C_m = \sum_i a_{m,i} \cdot P_i \tag{A2}$$

The change in attractiveness with price is then

$$-\frac{\partial C_m}{\partial P_i} = -\left(\frac{\partial a_{m,i}}{\partial P_i} \cdot P_i + a_{m,i} + (\text{secondary terms due to complex interactions})\right) \tag{A3}$$

This should be negative. If we ignore the secondary terms and use equation A1 for $a_{m,i}$ this implies

$$e^{-B_{m,i} \cdot P_i}(1 - B_{m,i} \cdot P_i) > \frac{-m_{m,i}}{A_{m,i}} \tag{A4}$$

for all feasible values for P_i. Since the right hand side is constant, we can examine the left hand side alone. It achieves a maximum with respect to P_i at

$$P_i = \frac{2}{B_{m,i}} \tag{A5}$$

At this point equation A3 will be close to a maximum (if the secondary effects are small). To avoid excessive positive values in equation A3 we should set

$$\frac{\partial a_{m,i}}{\partial P_i} \cdot P_i + a_{m,i} > 0 \tag{A6}$$

or with equation A1 and A5

$$m_{m,i} - A_{m,i} \cdot e^{-2} > 0 \tag{A7}$$

or

$$A_{m,i} < m_{m,i} \cdot e^2 \tag{A8}$$