ADVANCES IN SPATIAL SCIENCE

Francesca Pagliara
John Preston
David Simmonds

Residential Location Choice

Models and Applications

Springer

Advances in Spatial Science

For further volumes:
http://www.springer.com/series/3302

Francesca Pagliara • John Preston
David Simmonds
Editors

Residential Location Choice

Models and Applications

Springer

Editors
Assistant Professor Dr. Francesca Pagliara
University of Naples Federico II
Department of Transportation Engineering
Via Claudio 21, 80125 Naples, Italy
fpagliar@unina.it

Professor Dr. J. Preston
University of Southampton
Transportation Research Group
School of Civil Engineering
Highfield Campus, South
Southampton, SO171BJ, United Kingdom
j.m.preston@soton.ac.uk

Dr. D. Simmonds
David Simmonds Consultancy
Suite 14 Millers Yard, Mill Lane
Cambridge, CB2 1RQ, United Kingdom
david.simmonds@davidsimmonds.com

Advances in Spatial Science ISSN 1430-9602
ISBN 978-3-642-12787-8 e-ISBN 978-3-642-12788-5
DOI 10.1007/978-3-642-12788-5
Springer Heidelberg Dordrecht London New York

Library of Congress Control Number: 2010932612

Cover design: SPi Publisher Services

Printed on acid-free paper

Springer is part of Springer Science+Business Media (www.springer.com)

Contents

Contributors

J.E. Abraham HBA Specto Incorporated, 101 – 6th Avenue SW, Suite 1060, Calgary, AB, T2P 3P4, Canada, jea@hbaspecto.com

Theo Arentze Eindhoven University of Technology, Urban Planning Group, Eindhoven, The Netherlands, t.a.arentze@bwk.tue.nl

Pedro Donoso Department of Civil Engineering, University of Chile, Santiago, Chile

Jonas Eliasson Centre for Transport Studies, Royal Institute of Technology, Stockholm, Sweden, jonas.eliasson@abe.kth.se

Olga Feldman David Simmonds Consultancy, 14 Millers Yard, CB2 1RQ Cambridge, UK, OlgaFeldman@tfl.gov.uk

J.D. Hunt Department of Civil Engineering, University of Calgary, 2500 University Drive NW, Calgary, AB, T2N 2N4, Canada, jdhunt@ucalgary.ca

Jae Hong Kim School of Social Sciences, University of Ulsan, Ulsan, South Korea

Roger Mackett Centre for Transport Studies, University College, Gower street, London WC1E 6BT, UK

Francisco Martínez Department of Civil Engineering, University of Chile, Santiago, Chile, fmartine@ing.uchile.cl

Francesca Pagliara Department of Transportation Engineering, University of Naples Federico II, Naples, Italy, fpagliar@unina.it

John Preston Transportation Research Group, University of Southampton, Southampton, UK

Stephen H. Putman Department of City and Regional Planning, University of Pennsylvania, Philadelphia, PA, USA

Emma Richmond David Simmonds Consultancy, 14 Millers Yard, CB2 1RQ Cambridge, UK

David Simmonds David Simmonds Consultancy, 14 Millers Yard, CB2 1RQ Cambridge, UK

Harry Timmermans Eindhoven University of Technology, Urban Planning Group, Eindhoven, The Netherlands

Jan Veldhuisen Eindhoven University of Technology, Urban Planning Group, Eindhoven, The Netherlands

Paul Waddell Department of City and Regional Planning, University of California, Berkeley. Berkeley, CA, USA, waddell@berkeley.edu

T.J. Weidner Parsons Brinckerhoff Inc, 400 SW Sixth Avenue, Suite 802, Portland, OR, 97204, USA, weidner@pbworld.com

Alan Wilson Centre for Advanced Spatial Analysis, University College London, London, UK

Vassilis Zachariadis David Simmonds Consultancy, 14 Millers Yard, CB2 1RQ Cambridge, UK; Department of Architecture, University of Cambridge, 1-5 Scroope Terrace, Trumpington Street, Cambridge CB2 1PX, UK

The State-of-the-Art in Building Residential Location Models

Francesca Pagliara and Alan Wilson

Abstract This chapter provides an overview of the history of modelling residential location choice. Models of residential mobility typically have developed for illuminating the nature of location choice at different territorial scales or as part of an integrated model of land-use and transport. The latter tend to be more comprehensive in nature, though certain other investigations do consider interactions of location choice with other key decisions, such as work location.

Models presented in this book are described here briefly and are presented here according to three dimensions: theory and method, i.e. the modelling approach at the root of the model; categorisation of residential decision makers; and treatment of space, i.e. continuous, zoning or cells.

1 Introduction

Residential location modelling lies at the heart of one of the grand challenges of contemporary social science. More than 50% of the world's population now live in cities and, in different parts of the world. Effective planning demands a "What if?" forecasting capability and this can be achieved through the development of computer models. Since the elements of a city are highly interdependent, this in turn demands a comprehensive model of a city. Housing, where people live, how they choose their location – the elements of residential location modelling – is a critical element of this modelling task. Urban modelling represents a grand challenge because it can now be recognised as a generic task within the broader field that is now called *complexity science* – the science of understanding and modelling

F. Pagliara (✉)
Department of Transportation Engineering, University of Naples Federico II, Naples, Italy
e-mail: fpagliar@unina.it

A. Wilson
Centre for Advanced Spatial Analysis, University College London, London, UK

F. Pagliara et al. (eds.), *Residential Location Choice*, Advances in Spatial Science,
DOI 10.1007/978-3-642-12788-5_1,

nonlinear systems. The models presented in this book, therefore, are important for two reasons: first, they are key building blocks for urban models – and, indeed, in many instances, they are presented as components of general models; and secondly, they are exemplars of complexity science.

The models presented here represent the state-of-the-art. However, the ideas behind them have a long history. There are two main strands in this with a number of subsidiary elements, both with origins in the Nineteenth Century. There is a third early Twentieth Century strand which is essentially descriptive geography that we will note below but then not take further in relation to modelling. The first main strand is a focus on economics beginning with von Thunen's famous 1826 model of agricultural land use – "rings" of different kinds of crops around a market – and an associated theory of rent. The second is rooted in spatial interaction – mainly applied to migration or retailing and market areas – by analogy with gravity and hence the notion of "gravity models" – for example, see Carey (1858), Ravenstein (1885), Lill (1891) and in the Twentieth Century, Young (1924), Reilly (1929, 1931), Bossard (1932), Stewart (1942), Zipf (1946) and Iklé (1954). These models were all reviewed by Erlander and Stewart (1990). Neither of these strands, in the first instance, therefore, focused on residential location and structures. The subsidiary strand was contributed by geographers and sociologists: Burgess (1927) had a theory of rings but not based on bid rent – rather ecological notions of invasion and succession. Hoyt (1939) added sectoral differentiation while Harris and Ullman (1945) noted that expanding towns absorbed smaller towns and villages and that this added further polycentric structures. This in turn connects urban structure to the central place theories of Christaller (1933) and Losch (1940) but only indirectly to the theory of residential location.

Residential location modelling as we now know it dates back to the work of Alonso (1960, 1964) who laid the foundations for the economic analysis by applying von Thunen's key "bid rent" idea to residential location; and to Lowry (1964) who used spatial interaction principles in his *Model of metropolis*. Lowry used a very simple interaction model and earlier, Hansen (1959) had based a concept of "accessibility" on spatial interaction which was to play a role in many later models. Authors such as Carrothers (1956), Huff (1964) and Lakshmanan and Hansen (1965) had developed retail models which, again later, could be converted into improved residential location models. A variant on the interaction theme involved casting it in probabilistic form as in the work of Chapin and Weiss (1968). Alonso's model underpinned many future economic models while Lowry's initiated a host of comprehensive interaction-based models each of which had to have a residential location models. There were notable pioneering approaches, initially rooted in the big American transportation studies (e.g. Carroll 1955) which led to land-use transport models such as Penn-Jersey (Harris 1962). Many of the earliest of these are excerpted and described in Putman (1979). The models in this volume have their roots in one of these two strands. There are many histories of these developments: see for example, Batty (1976, 1994), de la Barra (1989), Wegener (1994) and Wilson (1998); Bertuglia et al. (1987) is particularly detailed. Eliasson and Mattson (2000), Iacono et al. (2008), Timmermans (2006) and

Wegener (2004) are more recent examples, with Hunt et al. (2005) providing a detailed review of the most recent large-scale operational modelling systems. Each of these provide routes to many more reviews.

These different perspectives have each been much developed and have to a large extent been integrated. An impetus for this integration came from Wilson's (1967, 1970) development of spatial interaction models on an entropy-maximising basis see also Senior and Wilson (1974). This facilitated the development of more complex models in which the hypotheses could be represented as constraints. There are detailed specifications of residential location models in Wilson (1970, 1974) and these models were tested by Clarke and Wilson (1985). Significant contributions came from Boyce (1978) and Kain (1987). However, the ongoing modelling task remains a formidable one! The system being modelled is immensely complicated and this means that researchers building empirical models have to compromise in various ways. This book offers an extensive range of empirical models and the various examples illustrate the range of choices that modellers have made to make their task feasible. It is interesting to summarise the dimensions of this complexity, how these relate to the roots we have identified and to note how the authors in this volume have responded to the challenges. We consider three main dimensions in turn: (1) theory and method; (2) the categorisation of residential decision makers; and (3) the treatment of space.

1.1 Theory and Method

Many factors can in principle influence residential choice. Lowry rooted his very simple model in the journey to work and the availability of employment. Access to services – such as "good" schools – is another interaction based element. Ways have to be found of characterising the type, quality and price of housing and this again can generate substantial arrays. Ideally, we need to capture the quality of different kinds of environment. Hypotheses on all these factors – and more – have to be incorporated in an underpinning the theory for the model. We have argued that the two starting points are the economic on the one hand – which has the advantage of generating surplus measures – spatial interaction modelling on the other. However, it can be argued, as noted earlier, that the two approaches can be integrated: the logit model and the entropy-maximising model are very closely related – see Wilson (2010) for a recent account of this relationship. One way or another, either the elements of a utility function have to be assembled and combined; or, equivalently, the components of the attractiveness functions in spatial interaction models. The spatial interaction formulation handles constraints very well and it is interesting in a number of these chapters that the importance of constraints has been recognised in other modelling approaches – for example with the development of constrained multinomial logit.

There is a particular theoretical issue associated with change – the dynamics. At any one time, the system of interest is almost certainly not in equilibrium and yet it is necessary at times to calibrate models assuming that it is. The representation of dynamics explicitly is made very complicated by the different "speeds" of the processes involved. There are elements of the population – ready to move – for whom the dynamics are "fast"; the developers are operating at the margin and so also can be considered to be part of a fast dynamics' process. But the whole system changes relatively slowly, though at times there will be phase transitions as whole neighbourhoods change character. It is particularly difficult to model changing land use. In terms of the spatial interaction formalism a method that can encompass phase changes and path dependence – cf. Arthur (1989) – was offered in a retail context by Harris and Wilson (1978) and articulated in a residential location context in Wilson (2000).

A final complication is that the effects of planning and zoning have to be allowed for.

1.2 Categorisation of Residential Decision Makers

The system is complicated by the variety of players. On the demand side, households can be characterised on a great variety of dimensions and this can create unmanageable arrays or model specifications for which there is no hope of assembling the data for effective calibration. On the supply side, the housing stock evolves slowly: developers can create new estates, householders can modify or extend their own properties. Housing is the great consumer of urban land and is in competition with other land uses and so "land" is a third major component of the system description. The finest level of detail which may be desirable produces unmanageable arrays and this leads to the possibility of using microsimulation as a method. This was pioneered by Orcutt (1957) and introduced in a spatial interaction context by Wilson and Pownall (1976). It is now in common use – see, for example, Clarke (1996) – and is used in the models in the Chapters on "Household Behaviour in the Oregon2 Model" by Hunt et al. and "A Microsimulation Model of Household Location" by Feldman et al.

1.3 Treatment of Space

Space can be treated as continuous or discrete – the latter case involving the creation of a zone system. In the limit, of course, a system described through a large number of very small zones mimics continuous space. Economic models, such as Alonso's, have tended to use the continuous representation. It is significant, however, that the translation of Alonso's work into discrete space by Herbert and Stevens (1960) was a significant precursor for the ongoing development of economic models. Interaction-based models nearly always use zone systems notwithstanding the work of Angel and Hyman (1976) in developing continuous space models. In practice, zone systems connect with available data more easily and the models are mathematically more tractable.

Cellular systems are a particular form of zone system and can then be connected to the literature on cellular automata.

When the elements of a residential location model are assembled, many of the components, more probably than for any other urban submodel, are themselves variables in other submodels: employment by location, services by location and transport costs for example. Such problems of interdependence are very difficult to handle outside the framework of a comprehensive model, and it is not surprising that most of the models presented here are developed within such a context.

All the models in this book use discrete zone systems, except for the Oregon model of the Chapter on "Household Behaviour in the Oregon2 Model" by Hunt et al., which uses a cellular system and the Edmonton model of the Chapter on "Stated Preference Examination of Factors Influencing Residential Attraction" by Hunt, which considers individual housing units but only in terms of demand and not in terms of supply or the UrbanSim model of the Chapter on "Modeling Residential Location in UrbanSim" by Waddell, which runs on dicrete zones, gridcells, or parcels depending on the model configuration.

2 Models Described in This Book

The varieties of residential location models presented in this book can be understood against this framework. The framework itself is summarised in Fig. 1.1 and in Table 1.1, the choices made in relation to the models in this book are indicated. In the rest of this chapter, we show how the contents of each chapter link to this framework. Table 1.2 reports the territorial geography of each model and its area of application.

The authors in this book represent a substantial proportion of the community that has the capability to build large urban models and to calibrate them empirically. It is fascinating to see the range of choices that have been made in the interests of feasibility. The reader will be able to tease out very easily the different ways in which the authors have characterised systems of interest and many have systematically reviewed the range of factors which could be incorporated in their models before almost inevitably, paring down their ideal lists.

There is a spin-off benefit from collecting these chapters together: to be able to see residential models developed for such a variety of national environments – covering virtually every continent. Table 1.2 reports the areas of application of the different models.

Table 1.1 summarises the different models reported into the classification shown in Fig. 1.1. The discrete choice/(possibly nested) logit model is much the most popular methodological base – used in the Chapters on "Stated Preference Examination of Factors Influencing Residential Attraction" by Hunt, "DRAM Residential Location and Land Use Model: Forty Years of Development and Application" by

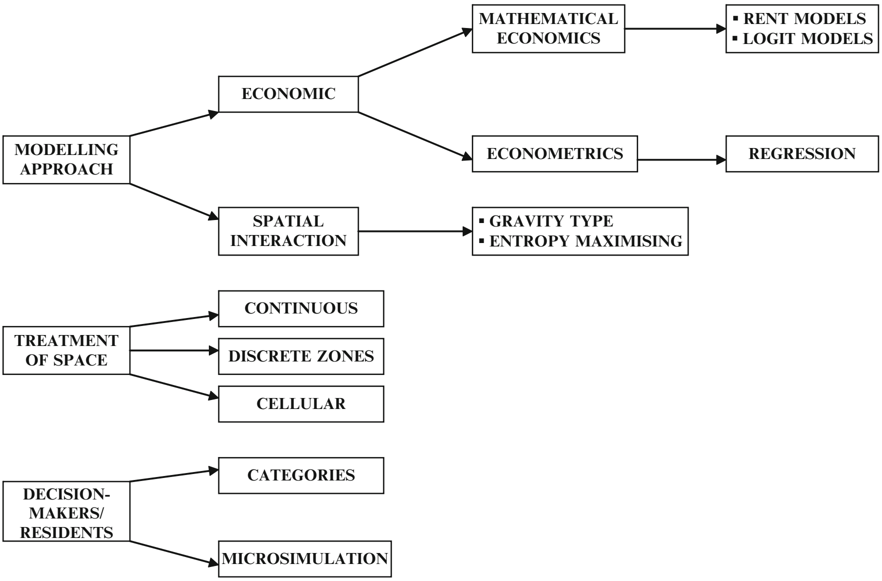


Fig. 1 Framework

Putman, "The Influence of Accessibility on Residential Location" by Eliasson, "Modeling Residential Location in UrbanSim" by Waddell and "The Residential Choice Module in the Albatross and Ramblas Model Systems" by Arentze et al. There are overlaps in these lists as models that start as spatial interaction models are converted into logit models for calibration purposes with a range of econometric methods deployed. The economic basis of the models figures strongly – consider for example the analysis of surplus in the Chapter on "DRAM Residential Location and Land Use Model: Forty Years of Development and Application" by Putman. The notion of bid rents still rates highly showing how Alonso's wonderful insights still play a major role (Chapters on "The MUSSA II Land Use Auction Equilibrium Model" by Martínez and Donoso, "The Impact of Transport Policy on Residential Location" by Pagliara et al. (2002), and "The Influence of Accessibility on Residential Location" by Eliasson). In Chapter "The Influence of Accessibility on Residential Location" by Eliasson and "The Residential Choice Module in the Albatross and Ramblas Model Systems" by Arentze et al., the economic analysis is explicitly linked to activity patterns. Many of the models emphasise their treatment of constraints in interesting ways (Chapters on "DRAM Residential Location and Land Use Model: Forty Years of Development and Application" by Putman, "The MUSSA II Land Use Auction Equilibrium Model" by Martínez and Donoso, "Modeling Residential Location in UrbanSim" by Waddell, and "The Residential Choice Module in the Albatross and Ramblas Model Systems" by Arentze et al.), including the introduction of constrained multinomial logit models. At this stage, the modelling of dynamics is typically on an incremental basis – essentially from one equilibrium to the next, period by period. In the Chapter on

Table 1.1 Models presented in the book set within the classification

Model – short name	Chapter	Author(s)	Modelling approach	Treatment of space	Decision-makers/residents
Edmonton	Stated Preference Examination of Factors Influencing Residential Attraction	Hunt	Mathematical economics – Logit models	Hypothetical individual residential units regarding demand only	Categories – Households divided into groups
DRAM	DRAM Residential Location and Land Use Model: 40 Years of Development and Application	Putman	Spatial interaction – Gravity type and Mathematical economics – Logit models	Discrete zones	Categories – Households divided into groups
DELTA	The DELTA Residential Location Model	Simmonds	Mathematical economics – Logit models	Discrete zones	Categories – Households disaggregated by composition, age, working status of working-age adults in the household, and the socio-economic group
MUSSA	The MUSSA II Land Use Auction Equilibrium Model	Martínez and Donoso	Mathematical economics – Constrained Logit and Rent models	Discrete zones	Categories – Households divided into groups by in-come, size and car ownership
Oxford	The Impact of Transport Policy on Residential Location	Pagliara et al.	Econometrics – Regression	Discrete zones	Categories – Households divided into income groups
TILT	The Influence of Accessibility on Residential Location	Eliasson	Mathematical economics – Logit models	Discrete zones	Categories – Households disaggregated into income groups
UrbanSim	Modeling Residential Location in UrbanSim	Waddell	Mathematical economics – Logit models	Dicrete zones, cells, or parcels	Categories – Households disaggregated into income, and size
Oregon2	Household Behaviour in the Oregon2 Model	Hunt et al.	Mathematical economics – Logit models	Cellular	Microsimulation
ALBATROSS RAMBLAS	The Residential Choice Module in the Albatross and Ramblas Model Systems	Arentze et al.	Mathematical economics – Logit models	Discrete zones	Microsimulation
SimDELTA	A Microsimulation Model of Household Location	Feldman et al.	Mathematical economics – Logit models	Discrete zones – ward level	Microsimulation

Table 1.2 Models areas of application

Model – short name	Geographical territory scale	Areas of application
Edmonton	Urban scale	Edmonton, Alberta
DRAM	Urban and metropolitan scale	Cities and metropolitan areas in USA
DELTA	Urban and regional scale	Cities and city regions in England, Scotland and New Zealand; national model of Scotland
MUSSA	Urban scale	Santiago, Chile
Oxford	Urban and regional scale	Oxfordshire
TILT	Regional scale	Stockholm region
UrbanSim	Metropolitan scale	Cities and metropolitan areas in USA and Western Europe
Oregon2	U.S.A. state scale	State of Oregon
ALBATROSS RAMBLAS	National scale	The Netherlands
SimDELTA	Urban and regional scale	South/West Yorkshire, England

"The DELTA Residential Location Model" by Simmonds, the notions of "mover pool" and "mobile population" are introduced in an interesting way. And as we noted earlier, a number of the models are rooted in a comprehensive framework and so represent interdependence. Logit models dominate model operationalisation – but it is interesting to see microsimulation beginning to appear significantly (the Chapters on "Modeling Residential Location in UrbanSim" by Waddell, "Household Behaviour in the Oregon2 Model" by Hunt et al., and "A Microsimulation Model of Household Location" by Feldman et al. – the later being a microsimulation version of the Chapter on "The DELTA Residential Location Model" by Simmonds model). Three of the chapters ("Stated Preference Examination of Factors Influencing Residential Attraction" by Hunt, "The Impact of Transport Policy on Residential Location" by Pagliara, and "The Residential Choice Module in the Albatross and Ramblas Model Systems" by Arentze et al.) use versions of stated preference methods to obtain their samples for model calibration. Finally, it will be noted that most of the models have been designed to contribute to the planning process and some engage explicitly with the zoning issue.

While the models reported here cover the main variety of residential location models, usually within a comprehensive framework, there are, of course, others that it has not been possible to include here. These are noted in the various reviews cited earlier. We note several of these approaches here to help complete the picture.

The MEPLAN model system (Echenique 2004), and the closely related TRANUS model system (de la Barra 2001) for developing integrated land use transport models include explicit representation of residential location. Work on these model systems, seeking a generalized representation for application in a range of different contexts, started in the 1980s, drawing on earlier models (Hunt and Simmonds 1993). These systems use a spatially disaggregated input–output structure to represent the behaviour of industry sectors and household categories and their interactions to simulate the spatial distributions of incremental production and floorspace rents arising from an initial allocation of exogenously generated "basic" components. The quantities of floorspace supply by type by geographic zone are adjusted in response to floorspace rents. The model system moves through time from one period to the next. In each period the distributions of incremental production are re-determined in response to changes in exogenous demand and floorspace. From one period to the next, the quantities of floorspace are adjusted in response to the changes in rents.

Households provide labor in response to industry demands at locations selected in response to wage rates, travel costs and the prices of key inputs, including residential space. In many specific applications, household expenditures are allocated among residential space, travel and other goods and services using a Cobb–Douglas function based on utility-maximizing assumptions. Travel costs are represented using a composite utility for the range of available mode alternatives between zones, these developed from a nodes-and-links representation of the available transportation supply. The list of practical applications of MEPLAN and TRANUS is extensive, including London, Napoli (Hunt 1994), Bilbao (Geraldes et al. 1978), Sacramento (Abraham and Hunt 1999), Sweden, Caracas and Mexico City and many others.

The PECAS model system for developing and applying spatial economic models also includes representation of elements of residential behaviour (Hunt and Abraham 2003). PECAS stands for "Production, Exchange, Consumption Allocation System". It is the name of a generalised framework and associated software system emerging since the year 2000.and now being used in a range of practical applications, including San Diego, Sacramento, Los Angeles, Atlanta, Baltimore and the State of California. (Hunt and Abraham 2005)

PECAS includes a computable general equilibrium structure for representing how "activities" (including industrial sectors, government and household categories) locate within the building space provided by developers and how these activities interact with each other at a given point in time. Flows of "commodities" (including goods, services, labour and space) going from production activity to consumption activity are determined according to technology options available to the activities and allocated from production location to exchange zone and from exchange zone to consumption location using an extended form of nested and additive logit model based on random utility theory (Abraham and Hunt 2007). Prices are determined for each commodity in each exchange zone in order to clear all markets. The actions of developers in the provision of the space (land and floorspace) consumed by activities in each zone – including the new development, demolition and re-development that occurs from one point in time to the next – are determined using a set of joint nested and continuous logit allocation models in response to relative prices, construction-related costs and zoning rules that specify allowable uses and intensities (Abraham and Hunt 2007). The resulting new quantities of space are used in the representation of the interactions among activities for the next point in time.

Households, as particular categories of activities, select residential locations, life-styles (as alternative technology options involving varying quantities of commodity production and consumption, including residential size and type), workplaces (as exchange locations for selling the labor "commodity"), the locations for other actions (as exchange locations for consuming other commodities) – all according to an extended form of nested and additive logit model whose parameters are calibrated for each household type as part of the development of a specific model. Households, as de facto developers when home owners, also determine whether or not to change the type and/or quantity of residential space from one point in time to the next.

Anas and Liu (2007) report the RELU-TRAN model. This is interesting as an economic model that is computationally challenging, interesting, and important but does not conform to linear, quadratic, or other standard nonlinear programming formulations. Rather, such models require the solution of highly nonlinear equations systems using non standard and innovative, iterative algorithms that exploit the special features of those equations. This is the approach that is used to design the RELU-TRAN algorithm. Numerical solution of models using iterative techniques has been a goal, though poorly practiced within the field of transportation and land-use modeling. Meanwhile, iterative numerical methods are gaining broader applicability within economics to solve a variety of problems.

RELU is a dynamic general equilibrium model of a metropolitan economy and its land use. It equilibrates floor space, land and labor markets, and the market for the products of industries, treating development (construction and demolition), spatial interindustry linkages, commuting, and discretionary travel. Mode choices and equilibrium congestion on the highway network are treated by unifying RELU with the TRAN algorithm of stochastic user equilibrium. The RELU-TRAN algorithm's performance for a stationary state is demonstrated for a prototype consisting of 4-building, 4-industry, 4-labor-type, 15-land-use-zone, 68-link-highway-network version of the Chicago MSA. The algorithm solves 656 equations in a special block-recursive convergent procedure by iterations nested within loops and loops within cycles. Runs show excellent and smooth convergence from different starting points, so that the number of loops within successive cycles continually decreases. The tests also imply a numerically ascertained unique stationary equilibrium solution of the unified model for the calibrated parameters.

RELU-TRAN is a spatially disaggregated, computable general equilibrium model based on microeconomic theory and in which economic activity is modeled at the level of fully interdependent model zones with a link-node transport network. It treats the stock of buildings in each model zone as changing slowly while other markets clear instantaneously. The metropolitan economy is treated as open in a number of ways. Consumers can locate their residences or jobs outside the metropolitan area, income can originate from outside and a part of assets within the area can be owned by outsiders, while firms can produce, in part, by paying for inputs located elsewhere. The model treats interactions between firms and consumers and among firms as purely pecuniary, which are sufficient to generate a pattern of spatial agglomerations.

Another example model is described in the work of Deal et al. (2005). They describe the Land-Use Evolution and impact Assessment Model (LEAM) which uses the STELLA/SME/GIS collaborative environment for the purpose of developing a Planning Support System (PSS) to generate and evaluate development patterns. It describes land-use changes across a landscape that result from the spatial and dynamic interaction among economic, ecological, and social systems in the region.

In the LEAM approach, groups or individuals who have substantive knowledge relating to a particular system develop and test separate models of that system. These contextual sub-models are linked and run simultaneously in each grid cell of a set of raster-based GIS map(s) to form the main framework of the dynamic spatial model (LEAM).

Inputs to the model use national land-use data sets (at 30 $\times$ 30 m resolution), census and economic data (readily available and transportable for application to multiple sites) along with variables relating to impact assessment sub-models (e.g. habitat, ecoregional inputs, water and energy inputs) to set model parameters. The products of LEAM model runs are analyses of a series of policy scenarios, presented as GIS maps or movies that show the transformation of the subject landscape as a product of policy related inputs. These dynamic visual outputs are beneficial for testing policy scenarios and raising concerns regarding the impacts of development, environmental degradation, or conflicting land-use policies. The final PSS tool will include a simple user interface and transportable data sets for application to multiple sites.

The economic model in LEAM (LEAMecon) forecasts changes in output, employment and income over time based on changes in the market, technology, productivity and other exogenous factors. The resulting economic trend is used as an input to a dynamic housing market simulation that then feeds into LEAM as residential land-use change. The agent-based housing model predicts actual houses built in a given year based on trends in the economy and anticipated demand by specific population cohorts. The combined economic and housing model serves as a prime driver of land-use change. Through LEAM, this work connects knowledge in regional science, housing markets, and spatial land-use analysis.

In the first substantive chapter “Stated Preference Examination of Factors Influencing Residential Attraction” by Hunt, a Stated Preference (SP) approach is used to develop a representation of household sensitivities to a range of both local and urban-level elements of residential locations. Each of a sample of respondents/residents in the population in Edmonton in Canada was asked to imagine moving the household to a new home location and to indicate preferences among hypothetical alternatives for this new location, with these alternatives described in terms of attributes related to the elements of interest, including housing type, mode specific travel times and costs for work and shopping, air quality, traffic noise, local street treatments, walking connectivity to local schools, and rent or taxes. The observations of choice behaviour thus obtained were then used to estimate logit choice models with utility function parameters indicating the sensitivities to these attributes. The results are indications about the influences on residential location and models incorporating representations of these influences.

In the Chapter on “DRAM Residential Location and Land Use Model: Forty Years of Development and Application”, the Putman DRAM model is presented as a component of the wider package ITLUP (Integrated Transportation and Land Use Package), which is arguably the first fully operational transportation-land use modelling software package. This has its origins in Putman (1983, 1991). It has now been applied in nearly 30 different metropolitan regions for public agency forecasting and policy analysis purposes. Designed on the Lowry framework, ITLUP offers a network representation that allowed for the incorporation of congested travel times in the distribution of activities. At the core of ITLUP are two allocation submodels: a household allocation submodel, which is DRAM, and an employment allocation submodel, EMPAL. Trip generation and distribution functions for the travel forecasting model are developed within DRAM, simultaneously with household location. Travel times from runs of the travel mode are fed forward to compute new activity distributions.

An interesting feature of DRAM is the location surplus measure, which defines the aggregate benefit households receive from the attributes of their chosen residential zone. The larger the value of location surplus, the more utility households receive from their choices of residential location. The surplus measures used in DRAM can be derived by using either of two different methods. Both methods produce the same location surplus measures and are based on the assumption that households attempt to maximize utility when choosing residential locations. For the first method, the DRAM model is interpreted as a multinomial logit model and the location surplus measure is found by calculating aggregate indirect utility. In the second approach, the location surplus measure is found by directly integrating the DRAM travel demand function.

The Chapter on "The DELTA Residential Location Model" by Simmonds describes the residential component of the DELTA land-use/economic modelling package. In its core, markets for residential and commercial real estate are represented, with transportation models linked into the overall model structure. The model system is divided into processes that represent spaces and those that represent activities. Processes dealing with activities include household formation and dissolution, employment growth or decline, location and property markets, and the employment status of individuals. The model system is designed to be run over a series of short steps of no more than 1 or 2 years. The main objective in creating this package has been that of creating a practical tool to forecast urban and regional change, and in particular to examine the expected impact of transport change; to provide a land-use/economic model which works in interaction with any appropriate transport model, and can therefore be used to extend relatively conventional transport models into land-use/transport interaction.

The location sub-model is both the "location and relocation sub-model", and the "property market sub-model". Mobile activities respond to changes in five variables: accessibility; quality of the local environment in general; quantity of housing; quality of housing; and the cost or utility of consumption, i.e. of spending income on housing, travel, and other goods and services.

DELTA is intended to be applied with a detailed classification of households reflecting household composition, age of household members, working status of working-age adults in the household, and the socio-economic group to which the household belongs. An important characteristic of the model is that only a proportion of households make residential choices in any one period. It is assumed that the main reasons for making a new residential choice are linked to change in one of the household classification variables, e.g. a change in the household's composition or in its work status. The households in the location model fall into two groups: “pool” households, which have no previous location within the area, and “mobile” households, which do have a previous location within the area modelled. Newly formed households and households resulting from existing households merging (e.g. singles forming couples) are assumed to make new location decisions and are counted as “pool” households. “Mobile” households are those which are undergoing other changes (mainly from couple with children onwards). In addition, a proportion of non-changing households is assumed to be “mobile” in each period. The numbers of “mobile” and “pool” households are initially calculated in the household transition model (which also finds and subtracts the numbers of households which have dissolved or migrated out of the modelled area altogether). The inter-area migration model is then applied, before the location model. The migration model predicts moves of households between areas within the modelled system: these households are subtracted from the “mobile” and “pool” numbers for the areas they leave, and added to the “pool” numbers for the areas into which they migrate. Households migrating from the rest of the world are also added to the “pool” numbers. The main location equations are weighted incremental logit functions, with slightly different forms for “pool” and for “mobile” households.

In the Chapter on “The MUSSA II Land Use Auction Equilibrium Model”, the MUSSA model is described by Martinez and Donoso. It is designed to forecast the expected location of agents, residents and firms, in an urban area. It presents an alternative framework for modelling land markets in transportation and land use models by adopting a modified version of the bid choice framework as it combines bid rent and discrete choice approaches to land markets by dealing simultaneously with both sides of an auction in an integrated framework. Real estate is allocated to the highest bidder by auction and market equilibrium is attained by the condition that all agents are located somewhere, and therefore supply satisfies demand. This auctioning process produces rents for each real estate in the market and simultaneously defines levels of satisfaction (benefits) to located agents at equilibrium. A discrete approach is followed for all units of demand and supply: households and firms are clustered into categories, while land is divided into zones and dwellings into types; the number of discrete units is defined by the modeller. Consumers’ agents, households and firms, are assumed rational and their idiosyncratic differences are modelled by a stochastic behaviour.

The place of MUSSA in the context of other land use models can be defined from a theoretical and historical perspective. A first generation of these models was designed under the assumption that agents locate as to minimize the travel cost to other activities, which may be called the maximum access model, where the transport system has a predominant role. Several models of this class were developed following either Alonso’s bid-rent approach or Lowry’s gravity – entropy – approach, or even a combination of these two. A second generation introduced market elements into the location problem by including rents and good prices, what we call the linear market model. Rents have been introduced in two ways, using a hedonic rent function based on average zone attraction indices, or by assuming the location options are quasi-unique so rents are the result of simulating an auction process known as the bid-auction approach. In this case, input–output tables have been used to incorporate spatially differentiated prices on goods. The third generation introduces an important amount of complexity into the model by incorporating an explicit representation of the direct interaction between agents decisions, that is the interaction that affect behaviour in addition to the price effects. These interactions describe the fact that location options are valued, by all agents and in a significant degree, by their built environment and the location pattern, usually called zone attributes.

A significant difference with other land use models is that in MUSSA the interaction between consumer agents – households and firms – is explicitly described in agents' behaviour and solved to attain equilibrium. This model was renamed as CUBE LAND with application currently being developed in several cities of USA, Europe and Asia.

The models developed for Oxford (Pagliara et al.) in the Chapter on "The Impact of Transport Policy on Residential Location" are not integrated into a formal package. The aim is that of assessing the extent to which transport impacts on residential location decisions and hence on house prices and that of evaluating the extent to which transport policy decisions (such as road user charging, work place parking levies, changes to fuel duties or the provision of light rapid transit systems) affect housing markets. This was achieved by undertaking two Stated Preference (SP) experiments in the Greater Oxford area divided into discrete zones, each with around 100 respondents corresponding to householders disaggregated into income groups. The aim was to determine the key transport and location factors that householders take into account when determining their residential location. It was intended that the choice models developed from the Stated Preference experiments would be used in conjunction with data on house prices to produce a bid choice model. However, price data was not available at a detailed enough level of spatial aggregation to permit calibration of an appropriate bid choice model. Instead, the SP data was used to develop an Hedonic Pricing (HP) model. Validation tests indicate that the HP model provides more reliable forecasts of house prices than the SP model. The HP model was used to provide preliminary forecasts of the impact of transport improvements on house prices in the Greater Oxford area.

In the Chapter on "The Influence of Accessibility on Residential Location", Eliasson describes the influence of accessibility on the household's location decision has been modeled through the use of the comprehensive TILT (Tool for Integrated analysis of Location and Travel) model. The main theoretical contribution is an elaborate specification of what it is meant meant by "accessibility" in this context. This is done by assuming that households, disaggregated into income groups, make a joint choice of location and activity pattern subject to income and time constraints. This activity pattern implies a stochastic travel pattern, the expected value of which is known at the time of location. The locational utility then consists of four parts: the indirect utility of income and time net of housing cost and expected total travel time and travel cost, the direct utility of the optimal activity pattern, the direct disutility of the expected travel pattern and the direct utility derived from location characteristics. The locational utility is then used in a discrete choice model for the choice of residential location.

In the Chapter on “Modeling Residential Location in UrbanSim” Waddell describes the residential component of UrbanSim, which is a microsimulation model of land markets, noted as the most widely used model today by Metropolitan Planning Organizations in the United States. It uses a flexible, modular structure to implement models that can be adapted to differtent geographic units such as zones, gridcells or parcels. The model system emphasizes clear behavioural realism, and attempts to avoid abstract modelling assumptions that are not reflected in observed behavior.

The UrbanSim model system contains model components representing household and employment relocation and location choices, and real estate development. The function of the household location choice model is straightforward, as is the data structure on which it operates. A list of households, generated using a synthetic population synthesizer, is represented in the base year database as a table with one row per household. Each household has a unique identifier, attributes such as number of persons, income, number of workers, presence of children, and a unique identifier for its location. As the model system proceeds in the first simulation year, the demographic transition model adds new households to the household table, providing their characteristics and a unique identifier, but not a location identifier. Then the household relocation model simulates the choices of certain households to move from their current location, and resets the location identifiers of the moving households to a null value. As a result of these two models, then, the household table contains some households that have moved into the region, and some that have been predicted to move within the region. These locating households are selected by identifying all households in the table that have a null location identifier. This is the set of households that the location choice model is applied to.

The framework for the household location choice model, like most of the models in the UrbanSim model system, is a standard choice model. Although more sophisticated choice model structures can be used, the most common in practice is the Multinomial Logit Model (MNL). The underlying logic of the model is that households that are in the market for a location take into consideration their own characteristics, such as income, and household size, and consider some sampling of available, vacant housing units and consider their price and characteristics such as density, age, and accessibility to employment and other opportunities. The relative attractiveness of these alternatives is measured by their utility. The choice model then proceeds to compute the probabilities of making a location choice from the available alternatives, defined as vacant housing units, given the preferences and budget constraints of locating households. Once location probabilities are computed, the predicted choices are simulated, using one of the available algorithms to reflect different assumptions regarding how the housing market clears. The model proceeds in steps. After loading the model specification and coefficients from input data, the model selects the agents that will be making a choice.

UrbanSim is implemented in the Open Platform for Urban Simulation (OPUS), and runs on multiple operating systems, using standard desktop or laptop computers.

Computational performance is efficient, with run times reported for parcel level model of San Francisco of 2–3 min per simulation year, using a full population for microsimulation.

In the Chapter on “Household Behaviour in the Oregon2 Model” the Oregon2 model is described by Hunt et al. It uses a set of seven connected modules representing different components of the full system, each running in turn for each year of simulation. Two of the modules concern elements of household behaviour. The household allocations module provides an agent-based microsimulation of each household and each person, simulating the transitions and choices made by these agents over 1 year. The Land Development Module provides a representation of space development using 30 m $\times$ 30 m grid cells covering the model area, microsimulating development transitions occurring in each cell over 1 year. It determines changes in developed space over time and in response to potential policy actions involving pricing, regulation and infrastructure in both transportation and land use. The Household Allocations (HA) Module provides a fully disaggregate representation using an agent-based microsimulation of each household and each person, simulating the transitions and choices made by these agents over the period of 1 year. The intent is to perform an endogenous determination of changes in social characteristics, so as to provide a more complete and consistent representation of demographic changes over time and in response to a wide range of potential policy actions involving pricing, regulation and infrastructure in both transportation and land use.

An initial population of households and household members for use in the simulation, with all attribute values assigned, is synthesized for the year 1990 using a sampling process that draws on a disaggregate sample of actual households and relevant marginal distributions from the Census.

The state-of-the-art in transportation and land-use modelling is defined by current research efforts aimed at building comprehensive microsimulation systems of urban areas, with representation at the level of individual agents (persons, households, firms, etc.) and simulations of the behaviour of the entire population of interest. The advantages of adopting such modelling approach for urban systems are that urban systems are dynamic, with a significant time element and components changing a different speeds. The behaviours of these systems are complex, with interacting agents, complex decision-making processes, and significant probabilistic elements. Closed-form mathematical and statistical representations of urban systems often induce large amounts of bias and lead to poor forecasts. Chapters on “The Residential Choice Module in the Albatross and Ramblas Model Systems” by Arentze et al. and “A Microsimulation Model of Household Location” by Feldman et al. deal with this issue. MUSSA, UrbanSim and Oregon2 present disaggregated households at a level of detail sufficient to operate them in a static microsimulation format, where a representative sample is used within a microanalytic framework for short run applications. However, for long run forecasts, the population should be synthesized or updated to represent the dynamics of individuals and the environments within which they make choices.

In the Chapter by Arentze et al., they describe the residential choice component in the Albatross and Ramblas model systems. Both models are primarily activity-based models of transport demand. Their prime goal is to predict activity-travel patterns and associated traffic flows. The distribution of residential land use, in terms of households and persons, is exogenously given in the case of Albatross. The spatial distribution of residential land use plays a double role in the simulation of activity-travel patterns in both models. First, both models assume the construction of a synthetic baseline population at the start of the simulation period. To that effect, the number of individuals and their values on a set of sociodemographics in each postal area are predicted, reflecting the spatial distribution of residential land use. This distribution influences the activity agendas and the spatial–temporal constraints underlying the models. This data can be exogenous to the models, implying that the relevant distributions should be based on an external model or data source and the creation of the synthetic population takes places at each simulation run. Secondly, residential land use is an integral part of the dynamics in the model systems. In this case, the aging and redistribution of the population, partly reflecting residential choice behaviour, is internal to the model system. In this case, a special sub-model or module predicts housing choice behaviour as a function of sociodemographics, characteristics of the available dwelling stock, characteristics of the transport network, and possibly activity agendas.

In the Chapter "A Microsimulation Model of Household Location" Feldman et al. describe the development of SimDELTA, which is a new microsimulation model of individual and household changes and choices within a land-use/transport interaction modelling structure as a development of the DELTA model. The microsimulation components explicitly model changes to members of the sample over time (rather than, as in many other microsimulation models, generating a separate sample for each modelled period of the forecast). The microsimulation modelling is carried out at ward level. The major strength of the model is naturally its disaggregate and dynamic nature, which means that the user can aggregate the output at any desired level of household or person characteristics, and that it is possible to trace individuals, households, jobs and dwellings over time so as to observe the modelled processes of change at a level of detail that is simply not possible in other types of model.

These chapters offer a very rich set of models. They are valuable in themselves and provide the foundation for the next generation of researchers working in this field. It is an effective representation of the present state-of-the-art.

References

Abraham JE, Hunt JD (1999) Policy analysis using the Sacramento MEPLAN land use transportation interaction model. Transp Res Rec 1685:199–208

Abraham JE, Hunt JD (2007) Random utility location/production/exchange choice, the additive logit model, and spatial choice microsimulations. Transp Res Rec 2003:1–6

Alonso W (1960) A theory of the urban land market. Pap Reg Sci Assoc 6:149–157
Alonso W (1964) Location and land use. Harvard University Press, Cambridge, MA
Anas A, Liu Y (2007) A regional economy, land use, and transportation model (RELU-TRANc_): formulation, algorithm design, and testing. J Reg Sci 3:415–455
Angel S, Hyman GM (1976) Urban fields. Pion, London
Arthur WB (1989) Increasing returns and lock-ins by historical events. Econ J 99:116
Batty M (1976) Urban modelling: algorithms, calibrations, predictions. Cambridge University Press, Cambridge
Batty M (1994) A chronicle of scientific planning: the Anglo-American modelling experience. J Am Inst Plann 60:7–16
Bertuglia CS, Leonardi G, Occelli S, Rabino GA, Tadei R, Wilson AG (eds) (1987) Urban systems: contemporary approaches to modelling. Croom Helm, London
Bossard JHS (1932) Residential propinquity as a factor in marriage selection. Am J Sociol 38:219–244
Boyce DE (1978) Equilibrium solution to combined urban residential location, modal choice and trip assignment models. In: Buhr W, Freidrich P (eds) Congestion among small regions. Nomos, Baden-Baden
Burgess EW (1927) The determinants of gradients in the growth of a city. Publ Am Sociol Soc 21:178–184
Carey HC (1858) Principles of social science. Lippincott, Philadelphia
Carroll JD (1955) Spatial interaction and the urban-metropolitan regional description. Pap Reg Sci Assoc 1:1–14
Carrothers GAP (1956) An historical review of the gravity and potential concepts of human interaction. J Am Inst Plann 22:94–102
Chapin FS, Weiss SF (1968) A probabilistic model of residential growth. Transp Res B 2: 375–390
Christaller W (1933) Die centralen Orte in Suddeutschland. Gustav Fischer, Jena. English edition: Christaller W (1967) Central places in Southern Germany (trans: Baskin CW). Prentice Hall, Englewood Cliffs, New Jersey
Clarke G (1996) Microsimulation for urban and regional policy analysis. Pion, London
Clarke M, Wilson AG (1985) The dynamics of urban spatial structure: the progress of a research programme. Trans Inst Br Geogr (New Series) 10:427–451
de la Barra T (1989) Integrated land use and transport modelling. Cambridge University Press, Cambridge
de la Barra T (2001) Integrated land use and transport modeling: the TRANUS experience. In: Brail RK, Klosterman RE (eds) Planning support systems: integrating geographic information systems, models and visualization tools. ESRI, Redlands CA, USA, pp 129–156
Deal L, Pallathucheril VG, Sun Z, Terstriep J, Hartel W (2005) LEAM – Technical Document Overview of the LEAM approach. University of Illinois at Urbana-Champaign. http://www.leam.uiuc.edu
Echenique M (2004) Econometric models of land use and transport, Chapter 12. In: Hensher D et al (eds) Handbook of transport geography and spatial systems, vol 5. Elsevier, Oxford, UK, pp 185–202
Eliasson J, Mattson L-G (2000) A model for the integrated analysis of household location and travel choices. Transp Res A 34:375–394
Erlander S, Stewart NF (1990) The gravity model in transportation analysis: theory and extensions. VSP, Utrecht
Geraldes P, Echenique M, Williams IN (1978) A spatial economic model for Bilbao. In: Proceedings of the PTRC Summer Annual Meeting, PTRC, London, UK
Hansen WG (1959) How accessibility shapes land use. J Am Inst Plann 25:73–76
Harris B (1962) Linear programming and the projection of land uses. Paper 20, Penn-Jersey transportation Study, Philadelphia
Harris CD, Ullman EL (1945) The nature of cities. Ann Am Acad Polit Soc Sci 242:7–17

Harris B, Wilson AG (1978) Equilibrium values and dynamics of attractiveness terms in production-constrained spatial-interaction models. Environ Plann, A 10:371–388

Herbert DJ, Stevens BH (1960) A model for the distribution of residential activity in an urban area. J Reg Sci 2:21–36

Hoyt H (1939) The structure and growth of residential neighbourhoods in American cities. Federal Housing Administration, Washington, D.C

Huff DL (1964) Defining and estimating a trading area. J Mark 28:34–38

Hunt JD (1994) Calibrating the Naples land use and transport model. Environ Plann 21B:569–590

Hunt JD, Abraham JE (2003) Design and application of the PECAS Land Use Modelling System. Proceedings of the 8th international conference on computers in urban planning and urban management, Sendai, Japan, May 2003, 16 pp

Hunt JD, Abraham JE (2005) Design and implementation of PECAS: a generalized system for the allocation of economic production, exchange and consumption quantities; Chapter 11. In: Foundations of integrated land-use and transportation models: assumptions and new conceptual frameworks. Elsevier, Oxford, UK, pp 217–238

Hunt JD, Simmonds DC (1993) Theory and application of an integrated land-use and transport modelling framework. Environ Plann 20B:221–244

Hunt JD, Kriger DS, Miller EJ (2005) Current operational urban land-use-transport modelling frameworks: A review. Trans Rev 25:1–48

Iacono M, Levinson D, El-Geneidy A (2008) Models of transportation and land use change: a guide to the territory. J Plan Lit. doi:10.1177/0885412207314010

Iklé FC (1954) Sociological relationship of traffic to population and distance. Traffic Q 8:123–136

Kain JF (1987) Computer simulation models of urban location. In: Mills ES (ed) Handbook of urban and regional economics, vol II. Elsevier, Amsterdam, pp 847–875

Lakshmanan TR, Hansen WG (1965) A retail market potential model. J Am Inst Plann 31:134–143

Lill E (1891) Das reisegesetz und seine anwendung auf den eisenbahnverkehr, Wien; cited in Erlander and Stewart (1990)

Losch A (1940) Die raumliche ordnung der wirtschaft. Gustav Fischer, Jena. English edition: The economics of location (trans: Woglam WH, Stolper WF(1954)). Yale University Press, New Haven, CT

Lowry IS (1964) A model of metropolis, RM-4035-RC. The Rand Corporation, Santa Monica, CA

Orcutt GH (1957) A new type of socio-economic system. Rev Econ Stat 58:773–797

Pagliara F, Preston J, Kim JH (2002) Modeling Residential Location Choice Behaviour. Proceedings of the European Transport Conference in 2004. Cambridge, UK

Putman SH (1979) Urban residential location models. Martinus Nijhoff, The Hague – Leiden, Holland

Putman SH (1983) Integrated urban models: policy analysis of transportation and land use. Pion, London

Putman SH (1991) Integrated urban models 2. New research and applications of optimization and dynamics. Pion, London

Ravenstein EG (1885) The laws of migration. J R Stat Soc 48:167–227

Reilly WJ (1929) Methods for the study of retail relationships, Bulletin No. 2944. University of Texas, Austin

Reilly WJ (1931) The law of retail gravitation. G. P. Putman and Sons, New York

Senior ML, Wilson AG (1974) Explorations and syntheses of linear programming and patial interaction models of residential location. Geogr Anal 6:209–238

Stewart JQ (1942) A measure of the influence of population at a distance. Sociometry 5:63–71

Timmermans HJP (2006) The saga of integrated land use and transport modelling: how many more dreams before we wake up? In: Axhausen K (ed) Moving through nets: the physical and social dimensions of travel. Elsevier, Oxford, pp 219–239

von Thunen JH (1826) Der isolierte staat in beziehung auf landwirtschaft und nationalokonomie, Gustav Fisher, Stuttgart. English edition: The isolated state (trans: Wartenburg CM (1966)), Oxford University Press, Oxford

Wegener M (1994) Operational urban models: the state of the art. J Am Inst Plann 60:17–29
Wegener M (2004) Overview of land-use transport models. In: Hensher D, Button K (eds) Transport geography and spatial systems. Pergamon/Elsevier Science, Kidlington, UK, pp 127–146
Wilson AG (1967) A statistical theory of spatial distribution models. Transp Res 1:253–269
Wilson AG (1970) Entropy in urban and regional modelling. Pion, London
Wilson AG (1974) Urban and regional models in geography and planning. Wiley, Chichester and New York
Wilson AG (1998) Land use-transport interaction models: past and future. J Trans Econ Policy 32:3–26
Wilson AG (2000) Complex spatial systems: the modelling foundations of urban and regional analysis. Prentice Hall, Harlow
Wilson AG (2010) Entropy in urban and regional modelling: retrospect and prospect. Geogr Anal, forthcoming
Wilson AG, Pownall CM (1976) A new representation of the urban system for modelling and for the study of micro-level interdependence. Area 8:256–264
Young EC (1924) The movement of farm population, Bulletin No. 246. Cornell Agricultural Experiment Station, Ithaca, NY
Zipf GK (1946) The P1P2/D hypothesis on the inter-city movement of persons. Am Sociol Rev 11:677–686

Stated Preference Examination of Factors Influencing Residential Attraction

J.D. Hunt

Abstract The City of Edmonton in Canada conducted a stated preference survey where over 1,200 respondents were asked to consider tradeoffs involving a wide range of elements of urban form and transportation, including mobility, air quality, traffic noise, treatment of neighbourhood streets, development densities and funding sources such as taxes. Respondents were to imagine moving to a new home location and to indicate preferences among hypothetical alternatives for this new location, with these alternatives described in terms of attributes related to the elements of interest. The observations of choice behaviour thus obtained were then used to estimate choice model parameters indicating the sensitivities to these attributes. As such, these parameter estimates provide indications of the relative importance of the corresponding elements and they also provide insights into the influences of the specific home location attributes considered. It is these insights into the influences of home location attributes that is of particular interest in this book presenting a collection of modelling treatments of household behaviour.

1 Introduction

The City of Edmonton in Canada developed a long-range transportation masterplan in the mid-1990s, encompassing a wide range of elements of urban form and transportation, including:

- Mobility
- Air quality
- Traffic noise
- Treatment of neighbourhood streets

J.D. Hunt
Department of Civil Engineering, University of Calgary, 2500 University Drive NW, Calgary, AB, T2N 2N4, Canada
e-mail: jdhunt@ucalgary.ca

F. Pagliara et al. (eds.), *Residential Location Choice*, Advances in Spatial Science,
DOI 10.1007/978-3-642-12788-5_2, © Springer-Verlag Berlin Heidelberg 2010

- Development densities and
- Funding sources such as taxes

The study described here was conducted as part of the development of the plan, in order to improve understanding of the relative importance placed on these elements by the population and thereby obtain some guidance concerning some of the tradeoffs to be made in the plan. The intention in this study specifically was not to consider what would be best for the population; but, rather, to consider the sensitivities of the population to a specific set of elements addressed in the plan.

A stated preference approach was used, where each of a sample of respondents in the population was asked to imagine moving to a new home location and to indicate preferences among hypothetical alternatives for this new location, with these alternatives described in terms of attributes related to the elements of interest. The observations of choice behaviour thus obtained were then used to estimate model parameters indicating the sensitivities to these attributes. As such, these parameter estimates provide indications of the relative importance of the corresponding elements, as required. But they also provide insights into the influences of the specific home location attributes considered. It is these insights into the influences of home location attributes that is of particular interest in this book presenting a collection of modelling treatments of household behaviour.

This chapter is organised into four sections after this introduction covering survey, analysis approach, results and conclusions.

The analysis approach involved the use of the standard logit model in the estimation of the indications of sensitivities. This particular form of mathematical model of discrete choice behaviour enjoys widespread use throughout the modelling of household behaviour. Consequently, it is appropriate to include in this first chapter (of a book about the modelling of household behaviour) a description of the standard logit model and its application – which is done within the section covering the analysis approach.

2 Survey

2.1 Survey Interview Design

The hypothetical new home location alternatives considered by respondents were described in terms of the following attributes:

- Auto drive time to work
- Operating cost for auto trip to work
- Parking cost for auto trip to work
- Transit ride time to work
- Walking distance to bus stop for trip to work
- One-way fare for transit trip to work

- Auto drive time to shopping
- Operating cost for auto trip to shopping
- Parking cost for auto trip to shopping
- Transit ride time to shopping
- Walking distance to bus stop for trip to shopping
- One-way fare for transit trip to shopping
- Housing type
- Change in housing taxes or rent relative to existing level
- Frequency of noticeably bad air quality
- Nature of traffic noise and disturbance arising from it
- Walking time to local elementary school
- Type of transportation facility to be crossed as part of walk to local elementary school and nature of any provision for that crossing and
- Type of street in front of dwelling and nature of any traffic calming measures in street

In order to remove other elements from consideration and thereby negate their potential impacts, each respondent was also told to assume that all other aspects of the alternative new home locations were the same as their existing home location. That is, for example, the respondent was to imagine that each alternative new home would have the same floor area and money value as the respondent's existing home location.

The descriptions of the hypothetical new home alternatives presented to respondents were developed by randomly varying the condition regarding each of the considered elements, with some degree of control on possible combinations of conditions in order to avoid inconsistent descriptions. That is, each specific hypothetical home location alternative was created by "bundling together" a randomly selected drive time to work, a randomly selected money cost for the auto trip to work, a randomly selected housing type, etc. An example of one restriction on the possible combinations was that a description could not have both a "collector road" in front of the dwelling and "none" for the level of traffic noise. Another was that parking charges remained the same across alternatives in a given interview, reflecting the invariance of work destinations.

Additional materials were presented to each respondent as part of the interview in an effort to establish a consistent understanding of the (sometimes fairly "jargony") terms being used. Separate single pages of point-form notes were used to indicate what was meant by the terms "shopping trip for groceries" and "noticeably bad air quality". Photographs (sometimes with additional point-form notes) were presented depicting the different housing types and indicating the meaning of local road, collector road, crosswalk, pedestrian bridge, block (as a distance measure), speed bump, and chicanes.

In each interview, the respondent was asked to participate in four separate stated preference "games", with four different hypothetical home location alternatives considered in each game. In a given game, the respondent was to establish his or her ranking of the alternatives in order from most to least preferred – which in general forced the respondent to make tradeoffs among the more and less preferable attributes and thereby provide data indicating the relative influences of these attributes The descriptions of the alternatives were printed on separate sheets of paper, one for each alternative. Figure 1 shows an example sheet presenting an alternative.

31401		2
Walk to local elem school:		**10 minutes** **using crosswalk on collector road**
Local air quality:		noticeably bad 1 day per year
Municipal taxes or rent:		**up \$125 per month (up \$1500 per year)**
Street in front of dwelling:		**local road**
Trip to work:	By car:	20 minutes in vehicle \$2.00 per day for parking \$0.75 for fuel & user charges (fuel taxes, road tolls)
	By transit:	45 minutes in vehicle with no transfers 4 block walk to bus stop \$1.25 fare each way
Trip to shop:	By car:	15 minutes in vehicle \$0.50 for parking \$2.00 for fuel & user charges (fuel taxes, road tolls)
	By transit:	30 minutes in vehicle with no transfers 1 block walk to bus stop \$1.25 fare each way
Traffic noise:		**constant faint hum**
Dwelling type:		single family

Fig. 1 Example sheet showing a hypothetical alternative. The respondent was asked to indicate preferences among four such alternatives for a new home location in each game by both ranking the alternatives in order of preference and also rating each alternative on a 0 to 10 scale

In the first three games the sets of four descriptions were selected randomly from the full set of available combinations of conditions. In the fourth game a fixed set of four pre-specified descriptions with specific themes was considered, with the same fixed set of four considered in all interviews. These four pre-specified descriptions embodied specific themes emphasizing mobility, money costs, the environment and children as areas of general concern and had the most attractive conditions for the relevant elements in each case.

In each game, the four sheets containing the descriptions of the alternatives were literally set out before the respondent and the descriptions were reviewed by the interviewer using the point-form notes and photographs as appropriate. The respondent was then left to consider and physically move around the sheets as part of the determination of the order of preference among the four alternatives. When the respondent finally settled on an order of preference, he or she was then asked to rate each hypothetical alternative on a whole number scale from 0 to 10, where 0 represents "terrible", 5 represents "neutral" and 10 represents "excellent". This acted to confirm the indicated order of preference: if any inconsistency arose between the indicated order of preference and the 0 to 10 scores, where an alternative with a higher ranking got a lower score, then the order of preference and the scoring were revisited until the inconsistency was eliminated. After the four games of ranking were completed, the respondent was then asked to indicate which one of the elements had the greatest influence in the ranking process.

It would be a very complex and demanding task to consider and compare conditions regarding 19 elements across four alternatives. A significant proportion of respondents might find such a task too difficult and either give up or simply provide incomplete responses, significantly reducing the accuracy of the results obtained. In anticipation of this potential difficulty, the number of elements varying in each of the first three games was reduced by organizing the elements into eight related groups and holding the conditions for four of these groups constant across all alternatives. The four groups of elements to be held constant in a given game were selected randomly as part of the development of the interview. This effectively eliminated these four groups of elements from consideration in that game, and thereby simplified the ranking task to a manageable level. All 19 elements were still printed on each sheet so that the full context was still indicated, and thus could influence the 0 to 10 rating scores. The four groups of varying elements were presented in bold print in order to allow them to be identified more easily.

The order of presentation of the elements could influence respondents, where elements at the top of a sheet might receive more attention than those further down because of respondent fatigue. In order to avoid this effect biasing the results, the order of presentation of both the groups of elements and the elements within the groups was selected randomly for each interview.

As part of the interview, the respondent was also asked to provide the following socioeconomic information:

- Nature of household tenure (own or rent)
- Taxes or rent paid

- AGE, gender, drivers license status, employment status, workplace location, frequency of travel to work and mode usually used for travel to work for each person in the household as appropriate
- Frequency and location of grocery shopping and mode usually used for travel to grocery shopping as appropriate
- Total number of private vehicles owned by household
- Present dwelling type
- Total household income and
- Nature of any mobility, sight/reading and/or hearing disabilities that might have influenced the responses provided

2.2 Conducting the Survey

A random listing of 6,000 households in Edmonton was selected from the publicly-available directory of residential telephone numbers. Several attempts were made to contact and recruit persons in these households for the survey. In the end, a total of 1,277 "face-to-face" interviews were successfully completed and the resulting information coded for analysis.

2.3 Resulting Sample as Representation of Population

In Fig. 2 the sample distribution of household sizes is compared with the corresponding population distribution – as observed in a 100% sample obtained in 1993 (City of Edmonton 1993). In Fig. 3 the sample distribution of dwelling types is similarly compared with the corresponding 1993 population distribution.

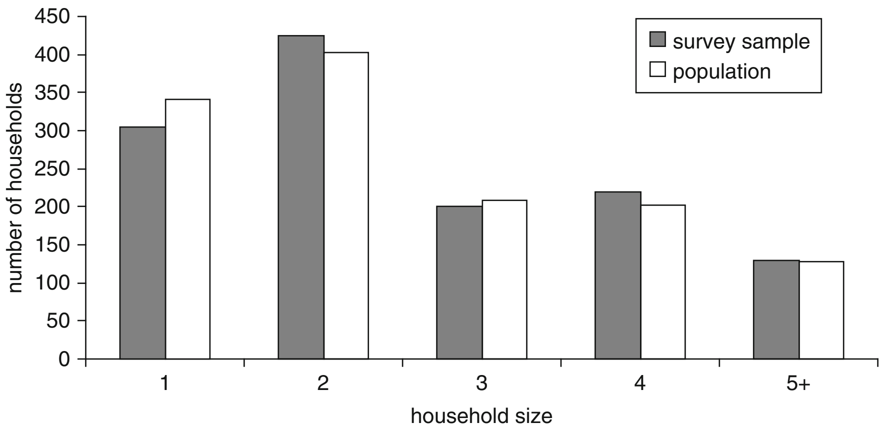


Fig. 2 Comparison of household size distributions in survey sample and population. The survey sample distribution is close enough to the population distribution in this regard to suggest that the sample can be taken to be representative of the typical Edmonton resident

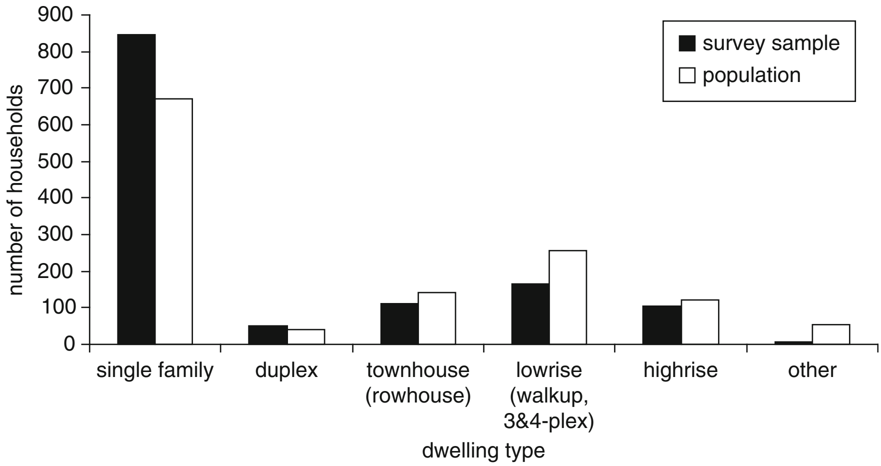


Fig. 3 Comparison of dwelling type distributions in survey sample and population. As above, the survey sample distribution is close enough to the population distribution in this regard to suggest that the sample can be taken to be representative of the typical Edmonton resident

These comparisons indicate that there is a reasonable match between the sample and the overall population regarding these distributions, with some relatively minor differences, suggesting that the sample can be taken to be reasonably representative of the population in discussions of the analysis results.

3 Analysis Approach

Indications of the influences of different attributes for specific groups of households were established by estimating standard logit models for those households using the observations obtained in the survey. The resulting parameters estimates for the logit model indicate the influences of the attributes. This is described below, covering the basic form of the logit model along with the estimation of the parameter values and the interpretation of the results.

3.1 Logit Model Form and Statistics

The logit model is a mathematical model that represents the behaviour of individuals trading off among the attributes of alternatives when selecting one alternative out of a set of available discrete alternatives (McFadden 1974). It has the following form for the choice situation considered here:

$$P_{i*} = \frac{exp(U_{i*})}{\sum_{i} exp(U_i)}, \tag{1}$$

where:

i	index representing new home location alternatives,
i*	a particular new home location alternative,
P_{i*}	probability that new home location alternative i* is selected,
U_i	utility value associated with new home location alternative i, expressed in (implied) hypothetical units called "utils".

The utility function that ascribes utility values to the new home location alternatives has the following general, linear form:

$$U_i = \phi_1 X_{1i} + \phi_2 X_{2i} + \dots + \phi_n X_{ni} + \dots, \tag{2}$$

where:

N	index representing attributes,
X_{ni}	value of attribute n for alternative i,
ϕ_n	utility function parameter associated with attribute n.

The mathematical form of the logit model is relatively simple and convenient to work with when using empirical data to estimate the values for the parameters, φ_n, in the utility function. The statistical properties of the resulting estimates are "well behaved" (McFadden 1974). Consequently, this formulation is a very attractive one for modelling choice behaviour and it continues to enjoy widespread use (McFadden 2007; Train 2003). Variations of this formulation providing more complex treatments have also been developed (Agresti 2007; Hensher and Green 2003; Koppelman 2006).

When values for the utility function parameters have been estimated, the relative influences of factors can be determined using ratios among the resulting coefficient values. For example, if for a given sample the parameter estimate associated with auto drive time is −2.00 utils per minute and the parameter estimate associated with transit ride time is −1.00 utils per minute then an increase in auto drive time is twice as onerous as an equal increase in transit ride time, indicating that auto drive time has double the impact and that a minute of auto drive time is worth 2 min of transit ride time for the typical household in the sample.

The significance of differences among estimates can be considered using standard t-statistics and t-ratios, with the t-ratio being the t-statistic for the estimate's difference from 0. When a t-statistic or t-ratio has a value greater than 1.96 in absolute magnitude, this indicates that there is a less than 5% chance that the associated difference is due to random effects only (Ang and Tang 1975), and the difference is said to be "significant".

The overall model goodness-of-fit can be considered using a goodness-of-fit index as follows (Ben-Akiva and Lerman 1985):

$$\rho^2(0) = 1 - \frac{L(*) - k}{L(0)}, \tag{3}$$

where:

K	number of coefficients in estimated model,
L(0)	log-likelihood for model with zeros for all coefficients,
L(*)	log-likelihood for model with estimated coefficients.

This $\rho^2(0)$ index is analogous to the R^2 statistic for linear regression in that it ranges from 0 to 1, with larger values indicating a better fit. It also takes into account the number of parameters used in the model, favouring more parsimonious model specifications (Ben-Akiva and Lerman 1985). Note that in this work specifically the k is omitted from the index as it is a constant across estimations.

A modified form of this goodness-of-fit index, that uses the model with just a full set of alternative specific constants as the point of comparison, is also often used. It has the following form:

$$\rho^2\left(C\right) = 1 - \frac{L(*) - k}{L(C)}, \tag{4}$$

where:

L(C)	log-likelihood for model with just a full set of alternative specific constants and zeros for all other coefficients.

This modified form provides further indication of the fit of the estimated model, in this case relative to a more informed point of comparison. The model with a full set of alternative specific constants uses the observed aggregate share selecting each alternative as the choice probability for the alternative. This is a more informed model that the one with zeros for all coefficients, which uses the inverse of the number of available alternatives as the choice probability for each alternative. Reports of estimation results sometimes include just one or both of $\rho^2(0)$ and $\rho^2(C)$. In cases where a model with a full set of alternative specific constants is not applicable, as is the case with the generic hypothetical alternatives considered in the first three games in the survey conducted for this work, then $\rho^2(0)$ is undefined and thus not reported. Other modifications to these indices are possible when the results of logit model estimations are reported: for example, the number of parameters, k, may not be included in the calculation of the index value, as is the case in this chapter.

The ALOGIT software package (ALOGIT 2007; Daly 1992) was used to estimate the parameters in this work. The "exploded logit" (Chapman and Staelin 1982) version of the estimation process was used, consistent with the use of a ranking process (rather than a single choice process) in the interviews. This "exploded logit" version makes full use of the ranking indications by attempting to predict the full ranking of the alternatives in an observation – not just the single, most preferred alternative.

The indication of the influence of a given attribute that is provided by the corresponding parameter estimate is a value that applies for the typical household for the sample being used. In this sense it is an average or "compromise" value that expresses a general tendency. The sensitivities of specific households in the sample to the attribute will most certainly differ from this compromise value. The variation in sensitivity may be very large – and specific households may feel very differently than the typical household. Therefore, the indications that are provided should only be applied in the consideration of general tendencies for large numbers of households.

4 Results

4.1 Presentation Format

Logit models were estimated for the entire sample and for various different sub-samples from the survey. The results for some of these estimations are discussed below, with the numerical values shown graphically in figures included in the text and also listed in tables included in an appendix.

4.2 All Households

The estimation results for the full sample of all households are shown in Fig. 4 and also listed in the first column of results in Table 1 of the Appendix.

Figure 4 shows the estimation results using a particular format, grouped and shaded by the categories of attributes considered. Working from left to right, these categories are: housing type, air quality, traffic noise, treatment of street in front of the dwelling, travel conditions to work, travel conditions to shop, and walking conditions to local elementary school. The bars for the groups show what the estimation results indicate would be the changes in utility occurring with the corresponding changes in attributes as either (a) increases (labelled "inc") or (b) switches from the relevant reference cases (the ones with zero values listed in the left-most position for each group).

In Fig. 4, utility values become more negative going up the page, indicating reductions in attractiveness. Thus, reading some examples from the figure: switching from "single family" to "duplex" for the dwelling type has about the same (or perhaps a slightly greater) negative impact on the attractiveness of an alternative as either changing the frequency of bad air from "never bad" to "bad 1 day per week" or increasing auto cost for travel to work by about \$6 (6 times the bar for "\$1 auto cost inc").

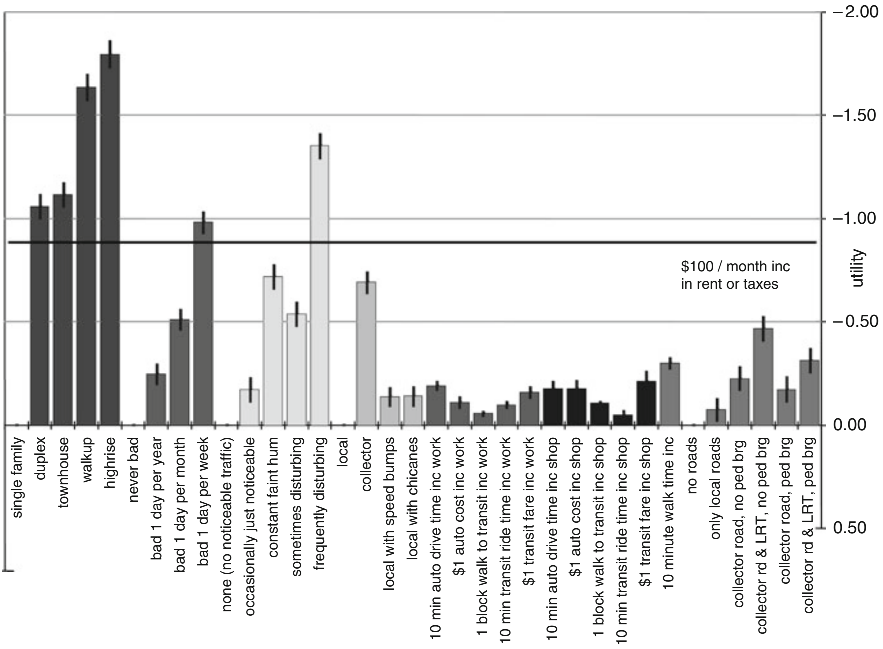


Fig. 4 Estimation results for sample of all households (1,277 households). The *bars* show the changes in utility that would arise with the indicated changes in attributes, with the change for a $100 per month increase in rent or taxes shown as a line to provide a reference. Changes in dwelling type away from single family have the greatest impact, followed by increases in traffic noise away from none, and then increases in the frequency of bad air quality away from never. Increases in travel times and costs to work and to shopping have comparatively modest impacts: the time to drive to work would have to increase by almost 60 min in order to have the same negative impact as a switch from single family to duplex dwelling type

The small thin vertical line at the top of each bar in Fig. 4 indicates the standard error for the statistical estimator of the value of the corresponding bar, thereby providing an indication of this aspect of the precision of the estimate. The impact of a $100 per month increase in taxes or rent is shown using a horizontal line rather than a bar in deference to its potential role as a basis for converting the shown changes in utility into corresponding changes in money amounts. As such, the particular format used for Fig. 4, along with the other similar figures included below, provides a useful graphical depiction of the indications regarding sensitivities provided by the estimation results.

Returning to the estimation results specifically, all the parameter estimates have signs consistent with expectations, and all are significantly different from zero except for those concerning (a) the sensitivity to the "only local roads to cross" condition regarding the walking trip to the local elementary school and (b) the set of three constants related to the four specific themes considered in the fourth game of each interview.

The ratio of the parameter estimate for auto time over the parameter estimate for auto cost is $0.1732 per minute. This is very close to previous estimates of the value of auto drive time in Edmonton: $0.1767 per minute from 1995 stated preference carpool use data (McMillan 1996) and $0.1962 per minute from 1991 revealed preference mode choice data (City of Edmonton 1991). This close match suggests that the value found here is a reasonably accurate estimate of the actual value of drive time to work for the population, which supports the idea that the respondents in this study are reacting to the hypothetical situations in much the same way Edmontonians have previously in other situations, which adds more general credence to the full range of results obtained in this study.

Similarly, the ratio of the parameter estimate for transit ride time over the parameter estimate for transit cost for the trip to work is $0.0619 per minute, which is reasonably consistent with the corresponding indications obtained in previous work.

The ratio of the parameter estimate for auto ride time to work over the parameter estimate for transit ride time to work is 1.94. This indicates that for the typical Edmonton resident a reduction in auto travel times to work has almost twice the impact on the attractiveness of residential locations as does an equivalent reduction in transit ride times to work. A value of 1.54 was obtained for this ratio in a previous study in Edmonton using revealed preference mode choice data (Hunt et al. 1998) and has been confirmed in later work (Hunt 2003). Values for this ratio of around 2 are not uncommon in work done elsewhere. Again, this reasonable match between behaviour in a real situation and behaviour in the corresponding hypothetical situation considered in this study adds credence to the results obtained in this study.

Respondents were told as part of the interview that it took 1 min to walk 1 block. This allows the estimation result for the walk to transit for the trip to work expressed in an amount per block to also be expressed as an amount per minute as required. The same applies for the result for the walk to transit for the trip to shopping.

The ratio of the parameter estimate for transit walk access to work over the parameter estimate for transit ride time to work is 2.83. This is somewhat higher than the value for this ratio found previously in Edmonton (Hunt 1990), but is similar to typical values found elsewhere (Ortúzar and Willumsen 1994). This may be due to sample error or some systematic distortion regarding the responses to transit walking distances in particular in this study. It is also possible that there are problems with the value found previously in Edmonton. In any case, this difference suggests that there should be some relatively greater concern about the results obtained here and in previous analysis in Edmonton regarding transit walk distances specifically.

The sensitivity to auto ride time for shopping is very similar to the corresponding sensitivity to auto ride time for work, with the work-related value being slightly higher. This is consistent with the idea that there may be additional time pressures arising because of concerns about required arrival times.

In contrast to the case with auto ride times, the sensitivity to auto cost for shopping is almost 60% higher in absolute magnitude than the sensitivity to auto cost for work, consistent with the notion that shopping travel is relatively more discretionary than work travel in general and therefore likely to exhibit higher cost elasticity.

The ratio of the parameter estimate for auto time over the parameter estimate for auto cost for shopping is \$0.1002 per minute. This is lower than the corresponding value of drive time for the trip to work and is in the range of similar values obtained work done previously, adding yet further credence to the range of results obtained here.

These results for the parameter estimates for auto ride time and auto money cost for both work and shopping together suggest that the difference in value of auto ride time between work and shopping arises because of a difference in the sensitivity to money costs, with the sensitivity to drive time remaining comparatively similar.

The ratio of the parameter estimate for transit walk access to shopping over the parameter estimate for transit ride time to shopping is 5.86. This is much higher than the corresponding value of 2.83 found for travel to work as indicated above. The type of shopping trip being considered likely had a substantial influence on this value in particular. In the interviews, respondents were told to consider a shopping trip for a week's worth of groceries for the household. Carrying these groceries for the walking component of the transit trip would make the corresponding walking distance or time relatively more onerous, contributing to the higher value obtained. This implies that the respondents made an effort to follow the interview instructions, at least in this respect, and that in general they behaved consistently, which is encouraging. It also means that the values associated with transit walk for shopping in particular must be interpreted carefully and applied elsewhere with caution.

The ratio of the parameter estimate for auto ride time over the parameter estimate for transit ride time for shopping is 3.74. This compares with a ratio of 1.94 for the analogous ratio for travel to work, suggesting that travel time to shopping has comparatively little impact on the attractiveness of home locations. This may arise because respondents are less likely to see transit as a viable alternative to auto for travel to shopping for a week's worth of groceries as defined for the shopping trip, and are therefore less concerned about the associated transit travel times. In any case, it follows that auto travel times for shopping have a much greater influence on the attractiveness of residential location than do transit travel time for shopping.

The ratio of the parameter estimate for transit ride time over the parameter estimate for transit cost for travel to shopping is \$0.0223 per minute. This is a relatively low value, consistent with expectations for this segment of travel, reflecting the comparatively higher sensitivity to money costs for shopping in general together with the reduced sensitivity to transit travel times for this situation in particular as discussed above.

The ratio of the parameter estimate for walk time to the local elementary school over the parameter estimate for walk time during the transit trip to work is 0.56. This lower sensitivity for the trip to school is expected to arise because trips to work are made by adults whereas trips to elementary school concern younger children with reduced time pressures generally. Respondents were told the walk time to the school was for a healthy adult, with the implication that the walk times for children would be somewhat greater and to the extent that the respondents took this into account it would act to make the sensitivity to walk time for the trip to school for children even lower than the value obtained in the estimation. Travel times for trips to the local elementary school may also have less influence generally because for many households there are few relevant work locations and many potential school locations spread throughout the region, reducing distances to schools generally and thereby making them less important when considering home locations. In any case, the travel times for trips to work have a much greater impact on the attractiveness of residential locations – for perhaps a range of reasons.

It is perhaps notable that substantial concerns about risks to children in the neighbourhood could have produced a high level of sensitivity to walk time for the trip to school – but that this did not happen. It may be that the typical Edmonton resident does not feel that such risks are significant in Edmonton generally. However, it may also mean that those respondents in households where such concerns are prevalent tend to drive children to school and therefore are relatively unconcerned about walking distances for the trip.

The estimation results for the "walking conditions to local elementary school" category of attributes indicate the influences (in terms of the differences in utility) associated with the types of roads or transportation corridors to be crossed and the facilities provided at the crossing points for the walking trip to the local elementary school. These differences in utility are expressed relative to the "no roads to cross" case, which has a fixed utility of zero.

The parameter estimates for all the other crossing situations are negative, indicating that any switch from having no roads to cross reduces the attractiveness of the residential location.

The t-ratio is only 1.4 for the parameter estimate for the "only local roads" condition for the walking trip to school. This suggests the need to cross some local roads, as opposed to no roads at all, on the way to school has a comparatively weak and dispersed impact on the attractiveness of residential locations.

The parameter estimate is -0.2251 utils for the "collector road, no pedestrian bridge" crossing on the walking trip to school. The ratio of this parameter estimate over the parameter estimate for walk time to school is approximately 7.5 min per trip, indicating that adding a collector road crossing on the walk to school has the same impact on the attractiveness of a home location as adding of 7.5 min to the walk itself. The ratio of this parameter estimate over the parameter estimate for money cost for the auto trip to work is approximately \$2.06 per trip. These ratios indicate there is a comparatively strong negative impact on the attractiveness of home locations, perhaps resulting from the perception of increased inconvenience and exposure to danger arising with the need to cross a collector road. The respondents were told that all such crossings without a pedestrian bridge were at painted crosswalks with pedestrian-actuated overhead flashing lights. It is expected that crossings without such facilities would be seen as even more onerous and have a correspondingly greater negative impact on the attractiveness of residential alternatives.

The ratio of the parameter estimate for "collector road with LRT, no pedestrian bridge" over the parameter estimate for walk time for the trip to school is approximately 15.6, indicating that adding a collector road with LRT crossing on the walk to school has the same impact as adding 15.6 min to the walk itself. This is an increase of 8.1 min, or more than double, relative to the collector road only condition; again, with a painted crosswalk. The implication is that the addition of an LRT crossing at grade on the walk trip to the local elementary school has a further large negative impact on the attractiveness of home locations, perhaps because of concerns about safety along with concerns about the impact of a light rail corridor on the sense of neighbourhood integrity.

The impacts of pedestrian bridges rather than painted crosswalks for these crossings on the walking trip to school are indicated with the parameter estimates for the "collector road, pedestrian bridge" and the "collector road with LRT, pedestrian bridge" conditions.

The addition of a pedestrian bridge reduces the negative impact of the collector road, but the differences in utility values are not statistically significant at the 5% level. The implication is that a pedestrian bridge on its own, when introduced over a collector road to be crossed on the walk to school, has a tendency to slightly reduce the negative impact on the attractiveness of residential locations, but that this is not a strong effect.

With the LRT included, the addition of a pedestrian bridge reduces the negative impact by an amount that is statistically significant at the 5% level, indicating that the introduction of a pedestrian bridge over a collector road with LRT does have a much more substantial positive impact. In terms of other factors, it has the same impact as a decrease in walking time of 5.2 min or a decrease in the money cost for auto travel to work of $1.41 per trip. However, because the parameter estimate for "collector road with LRT, pedestrian bridge", is still more negative than the corresponding parameter estimate for "collector road, no pedestrian bridge", the implication is that if LRT is added to a collector road corridor then the further introduction of a pedestrian bridge will not compensate for the addition of the LRT, at least with regard to the impact on the walk to the local elementary school.

The estimation results for the "dwelling type" group of parameters are expressed relative to the "single family" category, which has a fixed utility component of zero.

The parameter estimates for all the other dwelling types are negative, indicating that any switch from the single family dwelling type is undesirable, making the residential alternative less attractive.

The ratio of the parameter estimate for the "townhouse" category over the parameter estimate for auto travel time to work is approximately 58.9 min. This indicates that a switch from a single family dwelling to a townhouse is as onerous as an increase of almost 1 h in the daily trip to work by car for the typical Edmonton resident. The values for various other ratios concerning switches from "single family" can be established in a similar manner. For example, the value for the ratio for the switch from "single family" to "highrise" is nearly 95 min of auto travel time per trip to work.

These ratio values may seem high; but they are within the range of commuting drive times some households in larger cities are observed to accept in order to live in single family dwellings. The results obtained here indicate that the typical Edmontonian in the sample has similar attitudes and would behave the same way if forced to do so – but does not have to because single family alternatives are more generally available nearer to work. They are also very similar to analogous results obtained in a similar study performed in 1994 in Calgary, a comparable city also in Canada about 300 km south of Edmonton (Hunt 1994). Further, these ratios concern components of utility that apply with all other factors held constant. That is, everything else apart from housing category and auto travel time is assumed to remain the same. Various other unattractive aspects associated with being more than 90 min away from work, such as increased travel costs and reduced shopping and entertainment opportunities, can be expected to get mixed in and influence more intuitive attitudes concerning specific amounts of travel time, making these specific time amounts seem more onerous than they are on their own.

The components of utility for the various housing categories are consistently more negative with each increase in associated residential density, suggesting that increasing densities – at least at the micro-scale – act to reduce residential attractiveness. The changes in utility are relatively large, showing the largest changes of any of the groups of attributes, indicating that dwelling type, and all the various things associated with it, have a strong influence on housing preferences.

The estimation results for the "air quality" group of parameters are expressed relative to the "never bad" category, which has a fixed utility of zero. Not surprisingly, the parameter estimates for all the non-zero frequencies of bad air quality are negative and become more negative as the frequency increases.

The frequency of bad air going from "never bad" to "bad 1 day per week" has the same impact as an increase in municipal taxes of about $122 per month, or as an increase in auto drive time to work of about 52 min per trip.

It is perhaps somewhat surprising to find that there is such a strong feeling about the air being bad 1 day per year rather than never: it has the same impact on residential attractiveness as an increase in auto time to work of about 12.9 min per trip. To some extent it seems more reasonable that such a low frequency of bad air would have much the same effect as the air never being bad. But a similar level of sensitivity to the air being bad just 1 day a year was obtained in the above-mentioned study in Calgary (Hunt 1994). Perhaps the prevailing attitude is that if the air is worse than certain standards 1 day a year then it is likely to be less than ideal a number of times. In any case, there is a clear indication that the typical Edmontonian, like the typical Calgarian, is very sensitive to even very modest changes in air quality.

The estimation results for the "traffic noise" group of parameters are expressed relative to the "none" category, which has a fixed utility of zero. As expected, the parameter estimates are negative for all the cases where there is some degree of traffic noise.

The relative magnitudes of the parameter estimates for the traffic noise cases indicate that a constant faint hum of traffic has a somewhat more negative impact than does traffic noise that is "sometimes disturbing", and that "frequently distracting" traffic noise has the greatest negative impact. Going from "no noise" to "frequently distracting" traffic noise has the same impact on residential attractiveness as an increase in the drive time to work of about 71.4 min per day, and an even greater impact than does going from "air never bad" to "air bad 1 day a week".

At about the time the interviews were being conducted, a new truck route bylaw was being developed for the City of Edmonton. The additional media attention this generated on truck issues, including noise, may have caused some respondents to be somewhat more sensitive to traffic noise issues than is the case normally. The results indicated here should be interpreted accordingly.

Overall, the estimation results regarding the two environmental aspects considered indicate that these aspects can be very important to the typical Edmonton resident in the context of housing location depending on the extent of change involved. The relative levels of impact for the different levels of change considered seem entirely reasonable, which is seen to add further credence to the results.

The estimation results for the "street in front of dwelling" category of attributes are expressed relative to the "local" category (with no traffic calming treatments), which has a fixed utility of zero.

The parameter estimate for the "collector" category is -0.6885 utils, indicating that a change from "local" to "collector" in front of the dwelling has about the same negative impact on the attractiveness of home alternatives as a switch from "none" to "constant faint hum" regarding traffic noise. This impact of roads with a throughput function is perhaps a response to the associated higher levels of possibly faster traffic and corresponding range of adverse factors.

The parameter estimates for "local with speed bumps" and "local with chicanes" are both negative, indicating that the introduction of these treatments reduces the attractiveness of residential locations for Edmontonians. This is perhaps somewhat surprising. It was expected that traffic calming treatments acting to reduce traffic speeds would be appreciated on a local road in front of the dwelling and thus would act to increase the attractiveness of the location; with the general dislike of these treatments emerging only when encountering them while driving elsewhere. It may be that the results obtained here arose because respondents were unable to separate the appeal of such treatments in one context from the dislike of them in another, with the net effect being negative. It is also possible that some respondents may have been able to make such a separation, but saw the interview as an opportunity to send a message regarding these treatments generally. Some may even feel that these treatments alter traffic characteristics in such a way that traffic is no less and perhaps even more bothersome: the increased vehicle braking, accelerating and even bouncing that results from the treatments may be seen to create greater levels of disturbance that are worse than higher vehicle speeds. Additionally, some may feel that child safety concerns are not reduced much by these treatments – in fact there may be increased concerns about greater driver distraction and the potential for children to be hidden from drivers by horizontal obstructions with chicanes. Finally, ease of driving conditions in front of the dwelling may indeed be of more importance than reduced speeds even immediately in front of the dwelling at that point specifically. In any case, the results suggest that both speed bumps and chicanes on the street in front have a net impact reducing the attractiveness of residential locations.

The parameter estimate for an increase in rent or taxes of $100 is −0.8033 utils, which is a bit more than the impact of a switch from "none" to "constant faint hum" regarding traffic noise. Other comparisons can be made regarding the impact of tax increases, such as moving from a single family dwelling to a walkup, with a change in utility of −1.6340, is equivalent to increasing taxes by about $200.

As indicated above, the estimated values were not statistically significant for the three constants representing the differences in utility among the four pre-specified descriptions considered in the fourth game of each interview. There are only three such constants whose values are estimated because the fourth (arbitrarily selected a priori to be the "money costs" theme) is set to zero. Estimating these sorts of constants is a test of the ability of the full set of estimated parameters to explain the attitudes to a collection of elements altogether. The implication of these values not being statistically significant (that is, of the inability to reject the null hypothesis that they are zero) is that there is no net significance to what is omitted when the rest of the estimated parameters are used to explain attitudes to the collection of elements considered, which adds further credence to the results obtained.

4.3 Relative Influence of Elements for All Households

Based on the presentation in Fig. 4, the set of parameter estimates obtained for the full sample "loosely" suggests that housing type has the greatest influence on residential attraction, followed by traffic noise, air quality and municipal taxes. The nature of the road in front of the dwelling follows these. Then changes in auto travel times to work of 20 min or so – similar to the existing average travel time to work in Edmonton – and similar changes in walking times to school appear to be the next most influential after that, followed by other aspects of the walking trip to school and then by other components of the vehicle trips to work and to shopping.

These rankings of relative influence have an arbitrary component in that they are based on assumed values for various changes, such as the 20 min change in auto travel time used in the previous paragraph. The parameter estimations themselves (as listed in Table 1 of the Appendix) provide much less arbitrary indications – showing the rates at which various changes are valued and at which tradeoffs between different combinations of elements are made. That is, rather than merely indicate that air quality has more or less impact on residential attraction than does auto travel time to work or municipal taxes, these results indicate that completely avoiding bad air 1 day per month has the same impact as a decrease in auto travel time to work of about 27 min per trip or a decrease in municipal taxes of about $63 per month. This is less arbitrary and more precise. Altogether, these parameter estimates provide utility values for different housing alternatives that can be used to evaluate conditions more generally and also to forecast the actions of specific people and households in the selection of home locations.

Even with the above, it is still useful in some considerations to have some appreciation of the relative impact of each element independent of its specific condition. An alternative indication of the relative strength of influence of each element is available from the results via the t-ratio of the parameter estimate. The t-ratio provides a rough indication of the relative strength of the influence of the associated element on the utility for the alternative: if an element has a relatively strong influence it will tend to increase the magnitude of the associated parameter estimate relative to its standard error, making the t-ratio greater in absolute magnitude. Other factors also influence the t-ratio, which means that it is not an ideal indication of the relative strength of an element – but there is an overall general tendency for the t-ratio to be larger in absolute magnitude when the influence of the element is greater.

In addition, the interviews also included a direct question asking which element was most influential. Immediately after the ranking of hypothetical home locations, each respondent was asked to indicate which of the elements was the most influential in the ranking process.

Figure 5 is a “tornado diagram” that displays the t-ratio and the relative frequency of being indicated most influential side-by-side for each of the elements. The order of presentation from top to bottom is based on the relative frequency of being indicated most influential.

Generally, dwelling type, traffic noise, and municipal taxes or rent are the most important elements considered, followed closely by air quality and then by a second tier including walking time to school, auto ride time to work and the nature of the street in front of the dwelling. This order is broadly consistent with the order based on the interpretations of the estimation results as presented in Fig. 4 and discussed above. However, there are some differences towards the bottom of the diagram. Both of these measures become increasingly unreliable as relative importance decreases, which means that what can be said about the order of relative importance for the remaining elements below the nature of the street in front of dwelling is less definite, apart from the indication that aspects of the trip to work (beyond auto time) and the walking distance for transit to shopping are the most influential out of this lower tier. Perhaps most notable is the relatively weak influence of vehicle ride times to shopping.

4.4 Low Income Households

The estimation results for the sub-sample of 152 households with annual before-tax incomes less than $20,000 are shown in Fig. 6 (using the same general arrangement used in Fig. 4) and also listed in the second column of results in Table 1 of the Appendix.

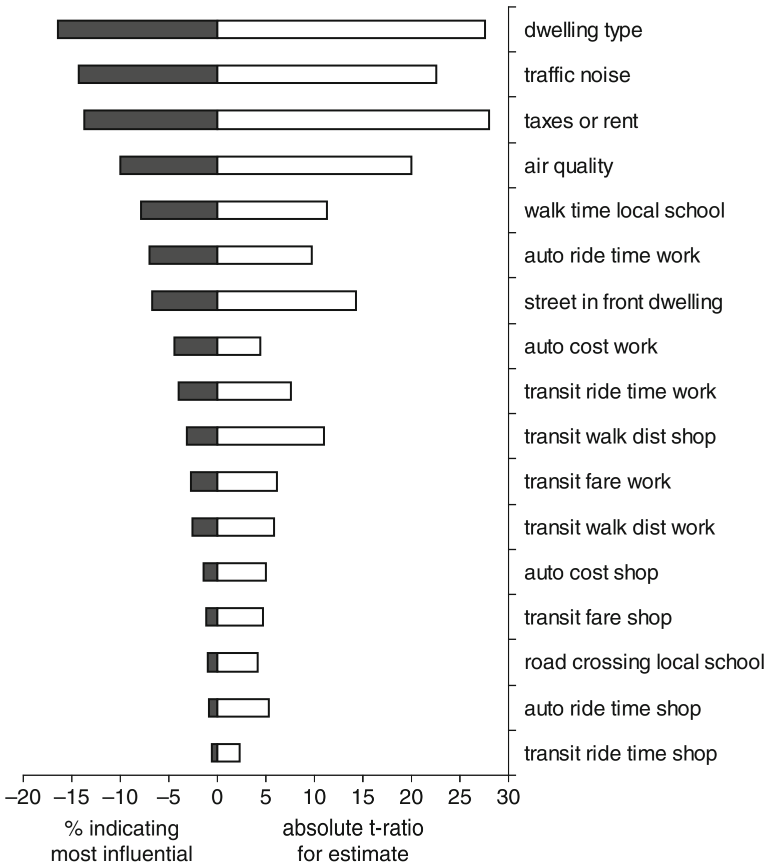


Fig. 5 Tornado diagram comparing t-ratio and percent most influential indications. The *bars* on either side of the centre line for each element show (*on the left*) the percent of interviews where the element was indicated to be the most influential in evaluating alternatives and (*on the right*) the absolute value of the t-ratio for the related parameter estimate for the element. Dwelling type, traffic noise, and municipal taxes or rent had the greatest influences, followed closely by air quality and then by a second tier including walking time to school, auto ride time to work and the nature of the street in front of the dwelling. Generally, travel conditions for transit and for trips to shopping had comparatively less impact on the attractiveness of home locations

The t-ratios are much smaller for these results than they are for the entire sample, consistent with the much smaller size of this sub-sample and resulting increased standard errors and associated reduced certainty in the parameter estimates.

The impact of auto travel time to work on the attractiveness of home locations is lower for this sub-sample. The ratio of the parameter estimate for auto time over the parameter estimate for auto cost for work is \$0.1363 per minute, lower than the corresponding value of \$0.1732 per minute found for the full sample. Obtaining a lower value for this sub-sample is consistent with the standard economic principle

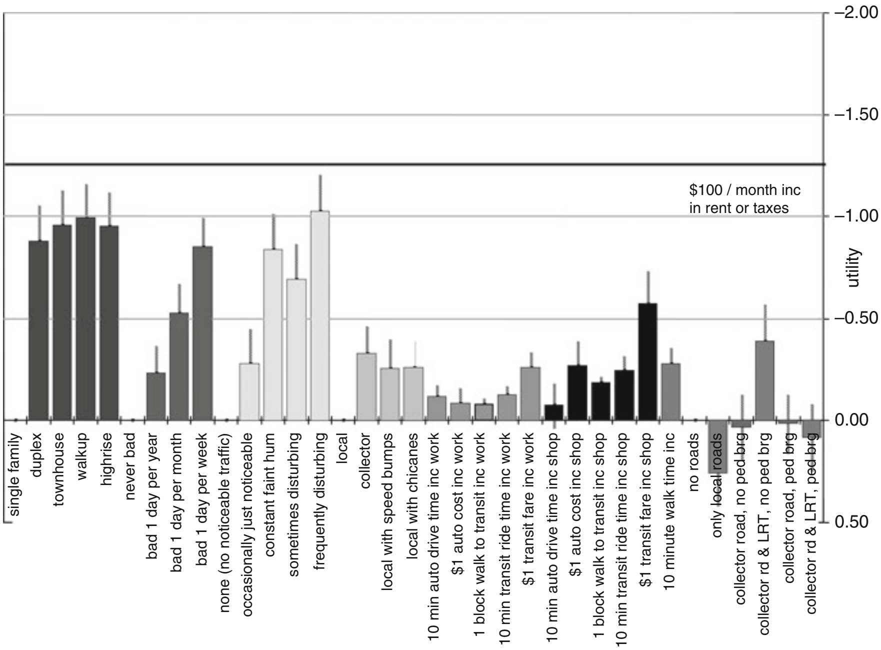


Fig. 6 Estimation results for sub-sample of low income households (152 households). The impacts of money amounts are all much greater than they are for all households, consistent with expectations. There is little difference among the impacts of dwelling types other than single family and road types other than local. The small sample size has likely influenced the results to some extent, with greater standard errors obtained throughout

that households with lower incomes tend to have lower values of time, all other things being equal, adding yet further credence to the range of results obtained here.

The impacts of the money costs of travelling to work and to shopping by transit on residential attraction are much greater for this sub-sample, as is the impact of municipal taxes or rent. This suggests that the typical lower income household tends to be much more sensitive to increases in money costs than does the typical "all-incomes" household, which is to be expected.

The impacts of changes to various housing types other than "single family" are reduced for this sub-sample and all roughly the same across the different types considered. For example, a change from "single family" to "walkup" acts to reduce residential attractiveness about as much as does an increase in transit cost to work of about $2.00 per trip. The differences in the parameter estimates among the dwelling types other than single family must be viewed cautiously, however, because the standard errors are much larger, indicating that the differences are fairly imprecise values.

The impacts of changes in air quality and traffic noise for this sub-sample are very broadly similar to what they are for the full sample. The nature of the street in front of the dwelling has somewhat less impact and the conditions for the walk to the local elementary school appear to have slightly greater impacts. The results concerning the impact of the need to cross only local roads is somewhat surprising: the positive estimate for "only local roads" implies that the need to cross some local roads is preferable to the ability to avoid all road crossings on the way to school. It is possible that this reflects concern about the personal security of children along more hidden paths behind houses rather than on the sidewalks of local streets, but it may also reflect a lack of knowledge generally within this sub-group of neighbourhood configurations that allow children to walk to school without crossing any roads. Again, because the standard errors are relatively larger, the values of the parameter estimates must be viewed with more caution.

4.5 High Income Households

The estimation results for the sub-sample of 113 households with annual before-tax incomes more than $100,000 are shown in Fig. 7 and also listed in the third column of results in Table 1 of the Appendix.

Again, the t-ratios for these results are much smaller than they are for the entire sample, consistent with the much smaller size of this sub-sample.

The sensitivity to auto travel time to work is greater for this sub-sample than it is for the full sample. The ratio of the parameter estimate for auto time over the parameter estimate for auto cost is $0.2092 per minute, higher than the corresponding value for the full sample, again consistent with the economic principle that those with higher incomes tend to have higher values of time.

The impact of municipal taxes or rent on the attractiveness of home locations is lower for this sub-sample than it is for the full sample, consistent with expectations.

The impact of the cost of travelling to work by transit is about the same for this sub-sample as it is for those in households with incomes less than $20,000 per year considered immediately above. This is somewhat surprising: it was expected that transit travel costs to work would have much less of an impact for this sub-sample.

The results concerning the cost of travelling to shopping by transit are also surprising: a positive parameter estimate was obtained for this sub-sample, indicating that the typical respondent in this group would prefer to pay more, not less, for transit to shopping. This is surprising, but it is consistent with the results obtained in the above-mentioned study done in Calgary (Hunt 1994) which also obtained a positive coefficient estimate for transit cost to work for the highest income group. It is hypothesized that these results do not arise because these households want to pay more themselves and thus find locations with greater transit costs more appealing; rather they arise because these households want to pay less in taxes. It is expected that these households tend not to use transit themselves, so they are not concerned about higher fares per se, but they are worried about the potential tax implications of higher transit deficits, which leads a significant number of the people in these households to feel that transit fares – non-work ones in particular – should be increased in order to reduce transit deficits. It bears noting that this hypothesis is

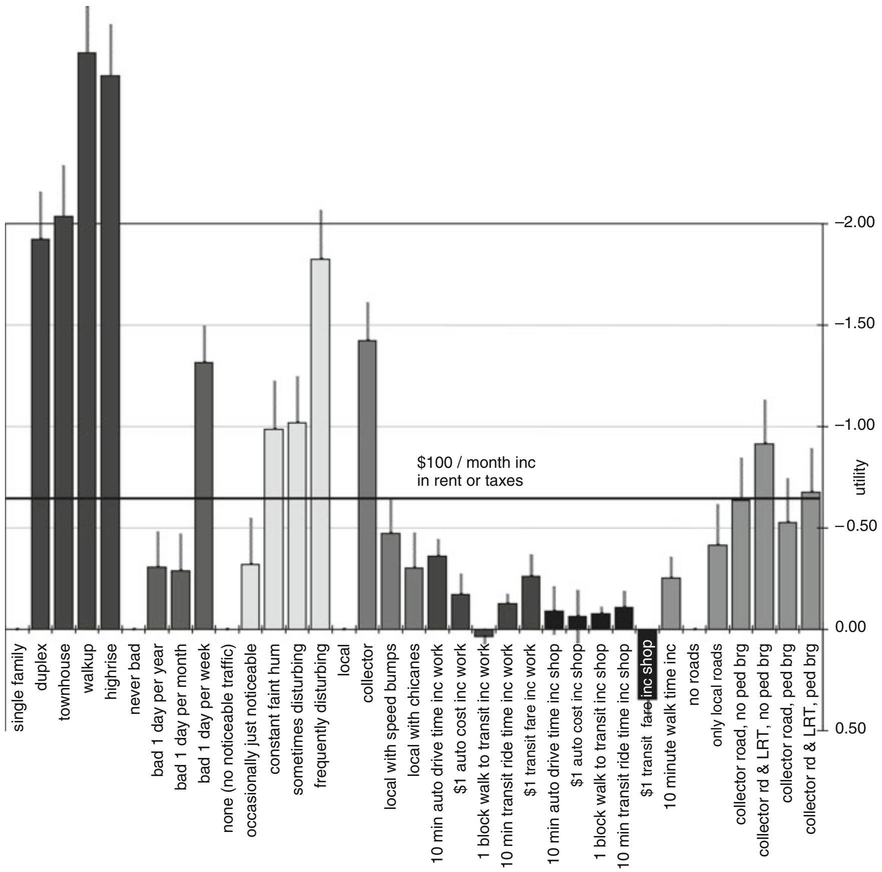


Fig. 7 Estimation results for sub-sample of high income households (113 households). The impacts of money amounts are all lower than they are for the full sample. The impacts of dwelling type, air quality, traffic noise and conditions for the walk to the local elementary school are all greater. The implied values of travel time (to work at least) are greater, consistent with expectations. The impact of the money cost for transit to shopping is positive, suggesting that increasing this fare will act to make housing locations more attractive. This unexpected result is perhaps caused by concerns within this sub-sample about the tax implications of higher transit deficits. Again, the small sample size has also likely influenced the results to some extent

highly speculative, and not based on information beyond that presented in the estimation results.

The impacts of changes to various dwelling types other than “single family” for this sub-sample are more pronounced than they are for the full sample, but they follow the same general pattern in that those types associated with higher densities have a comparatively greater impact. The effect on attractiveness of a change from “single family” to “highrise” is an exception: it is lower than the impact of a change from “single family” to “medium density” for this sub-sample whereas it is higher for the full sample. It may be that a significant proportion of those Edmontonians in this group imagine luxury-style (perhaps even penthouse) accommodation when considering the "highrise" housing type, which acts to reduce the overall negative evaluation of this dwelling type for this group.

The impacts of poorer traffic noise conditions and of reductions in air quality are generally greater for this sub-sample. This is consistent with the idea that those in this group will be relatively more concerned about these aspects and correspondingly less concerned about money costs overall. Still, the t-ratios for these parameter estimates are not as high as they are for the full sample, indicating that somewhat less confidence should be placed in the values found for these parameters.

4.6 Households with Children Under 18-Years Old

The estimation results for the sub-sample of 450 households with at least one member under 18-years old are shown in Fig. 8 and also listed in the fourth column of results in Table 1 of the Appendix.

The parameter estimates indicate that the attitudes and sensitivities for this sub-sample are fairly similar to those for the full sample.

One difference between this sub-sample and the full sample concerns the impacts of changes in dwelling type from "single family": types associated with higher residential densities have greater negative impacts for this sub-sample than they do for the full sample. The negative impact of a change from single family dwellings to highrises is particularly dramatic for this sub-sample, which is consistent with expectations for households with children.

Another difference concerns speed bumps and chicanes on the road in front of the dwelling. The parameter estimates for both are very near zero and highly insignificant in this case; whereas they are both negative and significant for the full sample. This suggests that there is some greater level of approval for these treatments in this sub-sample than is the case with the full sample – where those in households with children tend to see these treatments in a somewhat more positive light than do others – such that the net impact of these treatments on residential attractiveness is neutral rather than negative for the sub-sample.

4.7 Other Sub-samples of Households

The estimation results for various other sub-samples of households are shown in Tables 2–4 of the Appendix, as follows:

- In Table 2:

– Retired households, with all members over 65 years of age and not working (140 households)

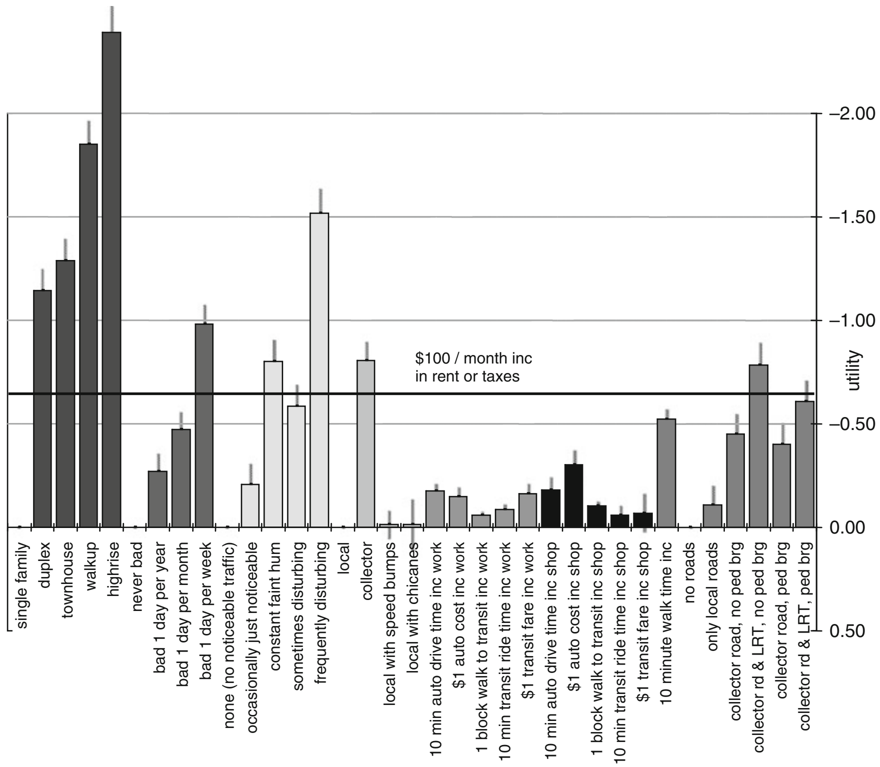


Fig. 8 Estimation results for sub-sample of households with children (450 households). The impacts of attributes are very similar to those for the full sample. The negative impacts of switching to dwelling types other than single family are greater and the impacts of traffic calming treatments are not as negative

- Unemployed households, with all members not working (262 households)
- Employed households, with at least one member working (1,015 households) and
- Households with no private vehicles (116 households)

• In Table 3:

- Households without children but not retired, with all members over 18-years old and at least one member working (687 households)
- Households located in the downtown area (176 households)
- Households located in inner city areas (736 households) and
- Households located in suburban areas (625 households)

• In Table 4:

- Households where alternative with environment emphasis most preferred in fourth game (453 households)

- Households where alternative with reduced money cost emphasis most preferred in fourth game (433 households)
- Households where alternative with children emphasis most preferred in fourth game (121 households) and
- Households where alternative with mobility emphasis most preferred in fourth game (270 households)

The estimation results for these other sub-samples are largely consistent with expectations.

For retired households and unemployed households the attributes related to travel to work have comparatively little impact on residential attractiveness, particularly for the auto mode. For households with no private vehicles and for retired households the attributes related to auto travel have relatively little impact – and the corresponding parameter estimates are all insignificant.

The consistency with results for households located in the downtown area tend to display attribute influences consistent with greater preferences and/or tolerances for downtown conditions, such as dramatically reduced aversions to dwelling types other than single family and much greater sensitivities to attributes of transit travel. The same applies broadly for inner city and suburban households. It should also be noted that there is some small amount of overlap between the definitions of inner city and suburban areas, so that the sum of the total number of households for the two corresponding sub-samples is greater than the number for the full sample.

The results for the sub-samples based on the preferences for particular emphases displayed in the fourth game are also consistent with the displayed preferences. But there is an element of circularity in these definitions – where the displayed preference that is the basis for the grouping into a particular sub-sample also contributes to the estimation results – which does reduce to some extent the strength of the indications provided in this regard, but only for these sub-samples.

Overall, the consistency with expectations displayed by these results is seen to provide further credence to the full set of results obtained.

5 Conclusions

5.1 Validity of Results

This study has successfully obtained valid indications of the impacts on residential attractiveness of a range of elements of urban form and transportation for various categories of Edmonton residents. The methods applied in the study avoided a number of anticipated difficulties as intended. The results are consistent with the findings of other work done in Edmonton and with standard economic principles. Furthermore, they match reasonable expectations. All this adds substantial credence to the results.

The sample of respondents interviewed appears to be a reasonable representation of the entire population of Edmontonians in spite of the biases inherent in its selection. There is a fair match between sample and population for some known characteristics and the attitudes of households are broadly consistent among most of the different sub-samples considered. Certainly there are differences among different sub-groups as indicated above – but the overall general pattern in the results remains somewhat the same, which can be expected to reduce the impact of differences between sample and population to some extent. Consequently, the results for the full sample are considered to be reasonably accurate indications of the attitudes for the typical Edmonton household, with the caveat that the sample is not 100% representative and therefore provides indications that could be slightly distorted. Certainly, it appears that the broad trends in the attitudes for the sample can be attributed to the population with reasonable confidence.

It should also be noted, as an additional caveat with regard to representation, that individuals were used to "speak" for households – to respond on behalf of their households. The potential differences between individuals and households and the associated issues regarding the representation of distributions for these two groups were not explored in this work.

As a final general caveat, in all cases the impacts on attractiveness indicated by the parameter estimates are for a "typical" individual as represented by the full sample or sub-sample considered. The sensitivities of specific individuals (or households) will most certainly differ from those determined for this typical individual. In fact, for example, some households even prefer highrise housing to single family housing. The values and tradeoff rates indicated here apply at the overall average level and should in general only be applied in consideration of broader tendencies for large numbers of Edmontonians.

5.2 Principal Findings

Out of the attributes of urban form and transportation considered, housing type, traffic noise and municipal taxes or rent have the greatest impacts on residential attractiveness for the typical household. The preferences for single family dwellings and little traffic noise in particular are very strong and consistent across almost all sub-groups. The typical household is willing to endure large increases in travel times and costs to work and shopping in order to stay in a single family dwelling or maintain low traffic noise, all other things being equal.

The typical household is also very concerned about municipal taxes and rents. The strength of this concern is broadly consistent across all groups considered. There is a much greater sensitivity to money paid in taxes than there is to money paid to travel to work or shopping.

Air quality is another attribute that has a relatively large impact on residential attractiveness. The desire for very low frequencies of bad air quality (according to government standards) is strong enough that the typical resident is willing to trade off substantial increases in municipal taxes and large increases in travel times to work in order to obtain these low frequencies of bad air quality, all other things being equal. This is very consistent across all sub-groups considered.

Walking times to the local elementary school and auto travel times to work have substantial impacts on the attractiveness of home locations. There is some variation in the levels of impacts of travel times and in the corresponding values of time, but overall the time spent travelling to school and work are some of the more influential attributes out of those considered. For the typical resident a decrease in travel time work of 1 min has the same impact on the utility of a residential location as a decrease in municipal taxes of $2.35 per month, all other things being equal. This provides a standard against which proposed transportation improvements could be assessed.

The nature of the street in front of the dwelling also has some reasonable impact on residential attractiveness for the typical resident. The desire for a local road rather than a collector is reasonably strong and consistent across all the sub-groups considered.

Auto travel times to work have more of an impact on household attractiveness than auto costs or transit travel times and costs to work for the typical resident. The sensitivities to auto travel times to work are roughly twice those to transit travel times to work. This means that the typical resident would get about twice the increase in residential utility with a reduction in auto travel times to work as he or she would with an equivalent reduction in transit travel times to work, all other things being equal. It also means that improvements to transit travel times to work must be roughly twice those to auto travel times to work in order to have the same impact on residential utility for the typical resident.

For the typical resident the money costs for travel to work or shopping have little impact on residential attractiveness relative to the other attributes elements that were considered.

As indicated above, the impacts of attributes on residential utility vary substantially across the population. Different sub-samples of households displayed different sensitivities, broadly consistent with expectations. The practical implication of this variation is that considerations based on a single set of values for the entire population are not going to respect this variation and thus are not going to be as accurate as considerations with different sets of values for different sub-groups of the population.

5.3 Further Work

Much further work could be done following on from the work reported here. Some of the possibilities considered most appropriate are outlined below.

The logit models and associated utility functions whose parameters have been estimated for the different categories of households could be used as models of residential location choice forming the basis of a residential allocation process in a land use model. These models would still have to be calibrated, adjusting the response characteristics and the aggregate shares to match known aggregate targets. This is because it is inappropriate to assume that the stated preference behaviour observed in this work provides valid indications of these aspects. But the trade-off rates among a wide range of elements for a variety of household types established in this work could be used directly.

The utility functions established in this work could also be used as the basis of a framework for policy evaluation. Utility values calculated using these functions could be used to evaluate the total change in satisfaction arising with changes regarding any one or more, and even all, of the housing related attributes considered – for the typical household or even for different types of households. This change in satisfaction can be expressed in dollar equivalents, thereby providing the essence of a cost-benefit analysis for policy evaluation. Such an analysis would include representation of the impacts regarding all of the attributes considered. For example, the impact on residential location utility of an increase in traffic noise in a given neighbourhood arising with the development of a new major road would be combined with the corresponding impact on utility arising with the improved travel times to work for all households. Thus, one of the criticisms of partial cost-benefit analysis is avoided in that these impacts are evaluated and included – at rates based on the indicated preferences of the typical household. Certainly, the changes in utility values for potential alternative plans would provide decision-makers with important numerical information regarding these alternatives, to be considered along with other information in deciding what to do for the future.

The sub-sample definitions based just on the socio-economic characteristics of the household members are most directly applicable in any further modelling or analysis work. Those sub-sample definitions based on the decisions of households, particularly those regarding location decisions, require the “answer” of the modelling process to be known before the appropriate model for the appropriate sub-group can be identified, which introduces a circularity that is avoided when definitions based just on the socio-economic characteristics of the household members are used.

The survey instrument and analysis used here could be similarly applied elsewhere, perhaps with modifications respecting differences in language and culture as appropriate. This would allow the development of a broader understanding of the sensitivities to attributes of urban form and transportation considered in this work – regarding how they vary in other settings and even across cultures, and thus the extent of any transferability in the associated parameter values. The use of a common survey instrument and treatment would help remove differences in results related to differences in conditions, and allow a more thorough consideration of the sensitivities themselves. The extent to which the within setting variations (among sub-groups) are greater than the between setting variations could also be considered – possibly helping develop a more complete understanding of the role of supply conditions and of sensitivities and attitudes in the development of urban areas.

The influences of other attributes could also be considered using the process used in this work. The set of attributes considered here is by no means exhaustive. A more complete understanding of the sensitivities of the typical household to a wider range of attributes would further inform both modelling and planning in various areas. For example, the impacts of road surface conditions, frequency of traffic lights along roadways, and safety at LRT stations, as examples within the transportation area, and of different forms of service charges and access to hospitals, libraries and social service programs, as examples within a larger context, could also be examined. Sensitivity to changes in the contribution to aggregate greenhouse gases consistent with the Kyoto Protocol could also be considered. This would provide indications of the tradeoff rates among such attributes If some of the attributes considered in this work are also considered in any such additional work, then a set of consistent tradeoff rates concerning the combined group of attributes can be developed. This would allow the combined impacts of any one or more, and even all, of the combined groups of attributes to be determined and compared on a consistent basis as described immediately above. Ultimately, this could lead to much more complete and consistent both modelling of behaviour regarding residential location and numerical-based consideration of the preferences of the population and of the quality of life being provided in the planning work that is done.

Acknowledgements The work described here was sponsored by The City of Edmonton. The preparation of this paper was supported by the Natural Science and Engineering Research Council of Canada and by the Institute for Advanced Policy Research at the University of Calgary. This description drew extensively from a previous paper (Hunt 2001). A more detailed description of this work is provided in the project report (Hunt 1996).

Appendix: Estimation Results

See Appendix Tables 1–4.

Table 1 Estimation results for Logit choice models for all households and selected sub-samples of households

Items		All households		Incomes below 20 K		Incomes above 100 K		Children under 18 years	
Parameters		Estimate	Abs t-ratio	Estimate	Abs t-ratio	Estimate	Abs t-ratio	Estimate	Abs t-ratio
Dwelling type	Single family	0	–	0	–	0	–	0	–
	Duplex	−1.0570	18.2	−0.8783	5.1	−1.9220	8.3	−1.1450	11.6
	Townhouse	−1.1130	19.1	−0.9566	5.6	−2.0350	8.2	−1.2880	12.8
	Walkup	−1.6340	26.2	−0.9895	5.9	−2.8410	12.6	−1.8520	17.0
	Highrise	−1.7940	27.8	−0.9514	5.8	−2.7290	11.0	−2.3930	19.5
Air quality	Never bad	0	–	0	–	0	–	0	–
	Bad 1 day per year	−0.2446	5.2	−0.2299	1.7	−0.3073	1.8	−0.2681	3.3
	Bad 1 day per month	−0.5092	10.6	−0.5264	3.7	−0.2879	1.6	−0.4741	5.8
	Bad 1 day per week	−0.9796	20.0	−0.8504	6.0	−1.3150	7.4	−0.9824	11.6
Traffic noise	None (no noticeable traffic noise)	0	–	0	–	0	–	0	–
	Occasionally just noticeable	−0.1694	2.9	−0.2758	1.6	−0.3192	1.4	−0.2051	2.1
	Constant faint hum	−0.7165	12.4	−0.8354	4.8	−0.9886	4.3	−0.8015	8.0
	Sometimes disturbing	−0.5348	9.3	−0.6934	4.1	−1.0180	4.5	−0.5852	5.8
	Frequently disturbing	−1.3500	22.2	−1.0250	5.8	−1.8230	7.6	−1.5200	14.0
Street in front of dwelling	Local	0	–	0	–	0	–	0	–
	Collector	−0.6885	14.1	−0.3307	2.5	−1.4250	7.7	−0.8073	9.3
	Local with speed bumps	−0.1352	3.0	−0.2570	1.8	−0.4735	2.9	−0.0126	0.2
	Local with chicanes	−0.1380	3.0	−0.2594	2.0	−0.3028	1.8	−0.0120	0.1
Taxes or rent	$100 per month increase	−0.8033	28.1	−1.2260	13.5	−0.6665	6.7	−0.6579	13.8
Travel to work	10 min auto drive time increase	−0.1890	9.6	−0.1170	2.1	−0.3613	4.4	−0.1756	5.2
	$1 auto cost increase	−0.1091	4.2	−0.0858	1.2	−0.1727	1.8	−0.1464	3.3
	1 block walk to transit increase	−0.0536	5.5	−0.0780	2.6	0.0340	1.0	−0.0568	3.4
	10 min transit ride time increase	−0.0974	7.2	−0.1253	3.1	−0.1246	2.6	−0.0876	3.9
	$1 transit fare increase	−0.1574	6.0	−0.2574	3.3	−0.2605	2.5	−0.1625	3.7
Travel to shopping	10 min auto drive time increase	−0.1746	4.9	−0.0747	0.7	−0.0918	0.8	−0.1804	3.0
	$1 auto cost increase	−0.1743	4.6	−0.2706	2.3	−0.0638	0.5	−0.3025	4.6
	1 block walk to transit increase	−0.1022	10.8	−0.1855	6.3	−0.0746	2.2	−0.1026	6.1

(*continued*)

Table 1 (continued)

Items		All households		Incomes below 20 K		Incomes above 100 K		Children under 18 years	
Parameters		Estimate	Abs t-ratio	Estimate	Abs t-ratio	Estimate	Abs t-ratio	Estimate	Abs t-ratio
	10 min transit ride time increase	−0.0467	2.1	−0.2458	3.6	−0.1077	1.4	−0.0601	1.6
	$1 transit fare increase	−0.2096	4.3	−0.5727	3.7	0.3610	2.2	−0.0692	0.8
Travel to school	10 min walk time increase	−0.2988	11.2	−0.2760	3.4	−0.2520	2.5	−0.5205	11.1
Crossing on walking trip to school	No roads	0	–	0	–	0	–	0	–
	Only local roads	−0.0740	1.4	0.2538	1.6	−0.4135	2.1	−0.1090	1.2
	Collector road, no ped bridge	−0.2251	4.0	0.0315	0.2	−0.6331	3.0	−0.4488	4.6
	Collector road with LRT, no ped bridge	−0.4666	8.1	−0.3897	2.2	−0.9123	4.3	−0.7852	7.8
	Collector road, ped bridge	−0.1723	3.0	0.0141	0.1	−0.5268	2.5	−0.4008	4.0
	Collector road with LRT, ped bridge	−0.3127	5.4	0.0802	0.5	−0.6774	3.2	−0.6090	6.2
Emphasis constants	Money cost	0	–	0	–	0	–	0	–
	Environment	0.2846	1.7	0.2145	0.5	0.1376	0.2	0.4744	1.7
	Mobility	−0.1027	0.6	−0.2146	0.5	−0.5058	0.8	0.1259	0.4
	Children	0.2074	1.3	0.8556	1.8	−0.3084	0.5	0.4072	1.5
Statistics									
Sample size	Number of households	1,277		152		113		450	
	Number of observations	15,315		1,824		1,356		5,394	
Goodness of fit	Initial likelihood	−16,224.0		−1,932.3		−1,436.5		−5,714.1	
	Final likelihood	−14,073.2		−1,649.1		−1,134.3		−4,885.3	
	$\rho^2(0)$	0.133		0.147		0.210		0.145	

Table 2 Estimation results for Logit choice models for selected sub-samples of households

Items		Retired		No workers		One or more workers		No private vehicles	
Parameters		Estimate	Abs t-ratio	Estimate	Abs t-ratio	Estimate	Abs t-ratio	Estimate	Abs t-ratio
Dwelling type	Single family	0	–	0	–	0	–	0	–
	Duplex	−0.8808	5.0	−0.9244	7.1	−1.1070	17.0	−0.6346	3.5
	Townhouse	−1.0350	5.8	−1.0150	7.9	−1.1650	17.6	−0.5522	2.8
	Walkup	−1.3180	6.5	−1.3460	9.7	−1.7380	24.6	−0.6712	3.5
	Highrise	−0.5076	3.0	−1.0750	8.2	−2.0150	26.9	−0.5052	2.8
Air quality	Never bad	0	–	0	–	0	–	0	–
	Bad 1 day per year	−0.1564	1.1	−0.1555	1.5	−0.2650	5.0	−0.4114	2.6
	Bad 1 day per month	−0.6202	4.3	−0.5377	5.2	−0.5110	9.4	−0.6081	3.7
	Bad 1 day per week	−0.9169	6.2	−1.0020	9.3	−0.9810	17.7	−1.0480	6.2
Traffic noise	None (no noticeable traffic noise)	0	–	0	–	0	–	0	–
	Occasionally just noticeable	−0.2288	1.4	−0.2649	2.1	−0.1335	2.1	−0.1549	0.9
	Constant faint hum	−0.5896	3.4	−0.7541	5.9	−0.7049	10.9	−0.5723	3.1
	Sometimes disturbing	−0.5223	3.0	−0.6133	4.7	−0.5151	8.0	−0.5636	3.0
	Frequently disturbing	−0.9789	5.6	−1.0330	7.9	−1.4450	20.9	−1.1870	5.9
Street in front of dwelling	Local	0	–	0	–	0	–	0	–
	Collector	−0.5932	4.2	−0.5512	5.2	−0.7360	13.3	−0.1613	1.0
	Local with speed bumps	−0.1478	1.1	−0.2305	2.3	−0.1155	2.2	−0.0058	0.0
	Local with chicanes	−0.3492	2.6	−0.2978	3.1	−0.1089	2.1	−0.0252	0.2
Taxes or rent	$100 per month increase	−0.9242	10.7	−1.0260	15.9	−0.7578	23.6	−1.0110	10.1
Travel to work	10 min auto drive time increase	0.0418	0.7	−0.0154	0.4	−0.2427	10.8	−0.0221	0.3
	$1 auto cost increase	−0.1197	1.5	−0.1128	2.1	−0.1057	3.6	0.0196	0.2
	1 block walk to transit increase	−0.0435	1.4	−0.0420	2.0	−0.0566	5.2	−0.0533	1.6
	10 min transit ride time increase	−0.0268	0.6	−0.0538	1.8	−0.1169	7.6	−0.1093	2.2
	$1 transit fare increase	−0.1727	2.0	−0.2142	3.6	−0.1474	5.0	−0.3336	3.7
Travel to shopping	10 min auto drive time increase	0.0679	0.6	−0.0812	1.0	−0.2000	5.0	−0.1069	0.9
	$1 auto cost increase	0.0891	0.7	−0.0600	0.7	−0.2008	4.7	−0.1782	1.4
	1 block walk to transit increase	−0.2740	8.2	−0.1913	8.5	−0.0824	7.8	−0.3055	8.5
	10 min transit ride time increase	−0.0304	0.4	−0.0571	1.2	−0.0416	1.7	−0.1507	2.0

(*continued*)

Table 2 (continued)

Items		Retired		No workers		One or more workers		No private vehicles	
Parameters		Estimate	Abs t-ratio	Estimate	Abs t-ratio	Estimate	Abs t-ratio	Estimate	Abs t-ratio
	$1 transit fare increase	−0.4562	2.9	−0.4656	4.2	−0.1459	2.7	−0.4974	2.8
Travel to school	10 min walk time increase	−0.1045	1.3	−0.1113	1.9	−0.3487	11.6	−0.3004	3.4
Crossing on walking trip to school	No roads	0	–	0	–	0	–	0	–
	Only local roads	0.1061	0.7	0.0273	0.2	−0.1075	1.8	−0.0999	0.6
	Collector road, no ped bridge	−0.1993	1.2	−0.1437	1.1	−0.2632	4.1	−0.0699	0.4
	Collector road with LRT, no ped bridge	−0.1871	1.1	−0.2780	2.1	−0.5222	8.0	−0.4749	2.4
	Collector road, ped bridge	−0.2584	1.5	−0.3011	2.4	−0.1505	2.3	−0.1555	0.8
	Collector road with LRT, ped bridge	−0.2000	1.1	−0.2460	1.8	−0.3503	5.4	−0.1755	0.9
Emphasis constants	Money cost	0	–	0	–	0	–	0	–
	Environment	0.2659	0.5	0.4152	1.2	0.2388	1.3	−0.0727	0.1
	Mobility	0.3365	0.7	0.3249	0.9	−0.2667	1.5	−0.1017	0.2
	Children	0.1466	0.3	0.5338	1.5	0.1100	0.6	0.0553	0.1
Statistics									
Sample size	Number of households	140		262		1,015		116	
	Number of observations	1,677		3,141		12,174		1,392	
Goodness of fit	Initial likelihood	−1,776.5		−3,327.4		−12,896.5		−1,474.0	
	Final likelihood	−1,542.9		−2,883.4		−11,095.1		−1,260.0	
	$\rho^2(0)$	0.131		0.133		0.140		0.145	

Table 3 Estimation results for Logit choice models for selected sub-samples of households

Items		No children not retired		Downtown located		Inner city located		Suburbs located	
Parameters		Estimate	Abs t-ratio	Estimate	Abs t-ratio	Estimate	Abs t-ratio	Estimate	Abs t-ratio
Dwelling type	Single family	0	–	0	–	0	–	0	–
	Duplex	−1.0730	13.5	−0.6198	3.7	−1.0660	14.2	−1.0740	11.4
	Townhouse	−1.0720	13.3	−0.1770	1.1	−1.0360	13.5	−1.2980	13.8
	Walkup	−1.6150	19.2	−0.9345	5.9	−1.5520	18.7	−1.7800	18.1
	Highrise	−1.7930	20.5	−0.3043	1.9	−1.5920	19.4	−2.1770	20.1
Air quality	Never bad	0	–	0	–	0	–	0	–
	Bad 1 day per year	−0.2562	4.0	−0.1625	1.3	−0.2766	4.5	−0.2427	3.2
	Bad 1 day per month	−0.5307	8.0	−0.6702	5.2	−0.6120	9.7	−0.3944	5.1
	Bad 1 day per week	−1.0320	15.4	−0.8823	6.5	−1.0290	16.1	−0.9842	12.3
Traffic noise	None (no noticeable traffic noise)	0	–	0	–	0	–	0	–
	Occasionally just noticeable	−0.1214	1.5	−0.0153	0.1	−0.2240	3.0	−0.0998	1.1
	Constant faint hum	−0.6977	8.9	−0.7601	5.0	−0.6924	9.1	−0.7686	8.3
	Sometimes disturbing	−0.5231	6.7	−0.5371	3.5	−0.6203	8.2	0.4040	4.4
	Frequently disturbing	−1.3850	16.9	−1.3430	8.6	−1.3420	16.8	−1.3940	14.3
Street in front of dwelling	Local	0	–	0	–	0	–	0	–
	Collector	−0.6778	10.2	−0.7240	5.2	−0.6785	10.4	−0.7125	9.3
	Local with speed bumps	−0.2041	3.3	−0.1967	1.5	−0.1559	2.5	−0.1128	1.6
	Local with chicanes	−0.1879	3.0	−0.2260	1.8	−0.1879	3.1	−0.0662	0.9
Taxes or rent	$100 per month increase	−0.9127	22.8	−0.8170	10.7	−0.8503	22.2	−0.7324	16.5
Travel to work	10 min auto drive time increase	−0.2503	9.2	−0.1968	3.3	−0.2061	8.2	−0.1749	5.3
	$1 auto cost increase	−0.0949	2.7	−0.1418	2.1	−0.1161	3.5	−0.0985	2.3
	1 block walk to transit increase	−0.0528	4.0	−0.0773	3.1	−0.0489	4.0	−0.0577	3.5
	10 min transit ride time increase	−0.1223	6.5	−0.1490	3.4	−0.0941	5.4	−0.0922	4.1
	$1 transit fare increase	−0.1551	4.3	−0.2884	3.9	−0.1591	4.7	−0.1521	3.5
Travel to shopping	10 min auto drive time increase	−0.2197	4.5	−0.0655	0.6	−0.2017	4.4	−0.1424	2.5
	$1 auto cost increase	−0.1347	2.6	−0.1738	1.8	−0.1119	2.2	−0.2447	4.1
	1 block walk to transit increase	−0.0746	5.9	−0.1228	5.0	−0.1085	8.6	−0.0964	6.4
	10 min transit ride time increase	−0.0431	1.4	−0.2095	3.2	−0.0704	2.5	−0.0152	0.4

(*continued*)

Table 3 (continued)

Items		No children not retired		Downtown located		Inner city located		Suburbs located	
Parameters		Estimate	Abs t-ratio	Estimate	Abs t-ratio	Estimate	Abs t-ratio	Estimate	Abs t-ratio
	$1 transit fare increase	−0.2401	3.7	−0.5145	4.0	−0.2634	4.1	−0.1127	1.5
Travel to school	10 min walk time increase	−0.2170	6.0	−0.3844	5.3	−0.2688	7.6	−0.3473	8.3
Crossing on walking trip to school	No roads	0	–	0	–	0	–	0	–
	Only local roads	−0.1086	1.5	−0.0509	0.4	−0.0447	0.6	−0.1576	1.9
	Collector road, no ped bridge	−0.1271	1.6	−0.2307	1.5	−0.2213	2.9	−0.2596	2.9
	Collector road with LRT, no ped bridge	−0.3500	4.4	−0.3600	2.4	−0.4777	6.2	−0.4893	5.4
	Collector road, ped bridge	−0.0298	0.4	−0.0438	0.3	−0.1274	1.7	−0.2589	2.8
	Collector road with LRT, ped bridge	−0.1840	2.3	−0.0697	0.5	−0.2662	3.5	−0.4211	4.6
Emphasis constants	Money cost	0	–	0	–	0	–	0	–
	Environment	0.1677	0.8	0.6687	1.5	0.0702	0.3	0.5476	2.1
	Mobility	−0.3496	1.6	0.4378	1.0	−0.0913	0.4	−0.1794	0.7
	Children	0.0658	0.3	0.8040	1.8	0.2141	1.0	0.1388	0.5
Statistics									
Sample size	Number of households	687		176		736		625	
	Number of observations	8,244		2,112		8,829		6,126	
Goodness of fit	Initial likelihood	−8,733.3		−2,237.3		−9,353.0		−6,489.6	
	Final likelihood	−7,450.2		−1,955.1		−8,118.0		−5,577.6	
	$\rho^2(0)$	0.147		0.126		0.132		0.141	

Table 4 Estimation results for Logit choice models for selected sub-samples of households

Items		Environment emphasis		Finances emphasis		Children emphasis		Mobility emphasis	
Parameters		Estimate	Abs t-ratio	Estimate	Abs t-ratio	Estimate	Abs t-ratio	Estimate	Abs t-ratio
Dwelling type	Single family	0	–	0	–	0	–	0	–
	Duplex	−1.3880	13.7	−1.1410	11.0	−1.2520	6.4	−0.5749	4.8
	Townhouse	−1.3510	13.0	−1.1460	11.3	−1.0970	5.6	−0.8295	6.8
	Walkup	−1.9100	17.2	−1.7070	15.9	−1.8370	8.8	−1.1330	8.5
	Highrise	−2.0380	18.0	−1.9340	17.3	−3.2020	10.4	−1.0450	8.3
Air quality	Never bad	0	–	0	–	0	–	0	–
	Bad 1 day per year	−0.3828	4.6	−0.1214	1.5	−0.4709	3.1	−0.2390	2.3
	Bad 1 day per month	−0.9812	11.0	−0.3191	4.0	−0.3043	1.9	−0.3780	3.8
	Bad 1 day per week	−1.5320	17.0	−0.7552	9.1	−0.8012	5.0	−0.7623	7.4
Traffic noise	None (no noticeable traffic noise)	0	–	0	–	0	–	0	–
	Occasionally just noticeable	−0.2156	2.1	−0.1602	1.6	−0.3055	1.5	−0.0712	0.6
	Constant faint hum	−1.0020	9.7	−0.6560	6.7	−0.8228	4.3	−0.4432	3.6
	Sometimes disturbing	−0.8275	8.1	−0.3692	3.8	−0.7978	4.2	−0.3338	2.7
	Frequently disturbing	−2.0330	18.0	−1.0340	10.3	−1.5790	7.5	−1.0260	8.1
Street in front of dwelling	Local	0	–	0	–	0	–	0	–
	Collector	−0.8689	10.4	−0.6462	7.5	−0.7689	4.5	−0.5752	5.6
	Local with speed bumps	−0.2753	3.5	−0.0930	1.2	0.1045	0.7	−0.1238	1.2
	Local with chicanes	−0.0957	1.2	−0.2545	3.2	−0.1217	0.8	−0.0739	0.7
Taxes or rent	$100 per month increase	−0.6660	14.0	−1.0570	20.0	−0.4710	5.2	−0.8631	14.0
Travel to work	10 min auto drive time increase	−0.1956	5.6	−0.1587	4.7	−0.2448	3.8	−0.2187	5.1
	$1 auto cost increase	−0.0653	1.4	−0.1841	4.1	−0.1100	1.3	−0.0778	1.4
	1 block walk to transit increase	−0.0280	1.7	−0.0434	2.5	−0.0480	1.6	−0.1154	5.5
	10 min transit ride time increase	−0.1221	5.0	−0.0721	3.1	−0.0077	0.2	−0.1633	5.5
	$1 transit fare increase	−0.0881	1.9	−0.2410	5.2	−0.0392	0.5	−0.1946	3.3
Travel to shopping	10 min auto drive time increase	−0.1518	2.4	−0.1755	2.8	0.0075	0.1	−0.2781	3.8
	$1 auto cost increase	−0.1405	2.2	−0.1669	2.5	−0.3645	2.8	−0.1973	2.4
	1 block walk to transit increase	−0.0695	4.3	−0.0979	6.0	−0.0638	1.8	−0.1793	8.4
	10 min transit ride time increase	−0.0108	0.3	−0.0782	2.1	−0.0043	0.1	−0.0564	1.2

(*continued*)

Table 4 (continued)

Items		Environment emphasis		Finances emphasis		Children emphasis		Mobility emphasis	
Parameters		Estimate	Abs t-ratio	Estimate	Abs t-ratio	Estimate	Abs t-ratio	Estimate	Abs t-ratio
	$1 transit fare increase	−0.1460	1.7	−0.2272	2.7	−0.0824	0.4	−0.2757	2.7
Travel to school	10 min walk time increase	−0.2231	4.9	−0.2874	6.3	−0.5572	6.2	−0.3633	6.2
Crossing on walking trip to school	No roads	0	–	0	–	0	–	0	–
	Only local roads	−0.1613	1.8	−0.0280	0.3	−0.2996	1.6	0.0472	0.4
	Collector road, no ped bridge	−0.2784	2.9	−0.1058	1.1	−0.9699	5.0	−0.0590	0.5
	Collector road with LRT, no ped bridge	−0.6001	6.0	−0.2636	2.6	−1.3160	6.4	−0.2956	2.4
	Collector road, ped bridge	−0.3898	3.8	0.0309	0.3	−0.7464	3.7	−0.0052	0.0
	Collector road with LRT, ped bridge	−0.3776	3.9	−0.1571	1.5	−0.8532	4.3	−0.2386	1.9
Emphasis constants	Money cost	–	–	–	–	–	–	–	–
	Environment	–	–	–	–	–	–	–	–
	Mobility	–	–	–	–	–	–	–	–
	Children	–	–	–	–	–	–	–	–
Statistics									
Sample size	Number of households	453		433		121		270	
	Number of observations	4,074		3,891		1,089		2,430	
Goodness of fit	Initial likelihood	−4,315.8		−4,121.9		−1,153.6		−2,574.2	
	Final likelihood	−3,430.6		−3,425.1		−922.5		−2,227.1	
	$\rho^2(0)$	0.205		0.169		0.200		0.135	

References

Agresti A (2007) Multicategory logit models, Chap. 6. In: An introduction to categorical data analysis, 2nd edn. Wiley, New York

ALOGIT (2007) Alogit Software and Analysis Ltd, available at the website: www.alogit.com. Accessed 2007

Ang AHS, Tang WH (1975) Probability concepts in engineering planning and design – vol 1: Basic principles. Wiley, Toronto

Ben-Akiva ME, Lerman SR (1985) Discrete choice analysis: theory and application to travel demand. MIT, Cambridge, MA

Chapman RG, Staelin R (1982) Exploiting rank ordered choice set data within the stochastic utility model. J Market Res 19:288–301

City of Edmonton (1991) Calibration of a nested Logit AM peak hour mode split model. Transportation Department, The City of Edmonton, Edmonton AB, Canada

City of Edmonton (1993) Report of the Edmonton Civic Census for 1993. The City of Edmonton, Edmonton AB, Canada

Daly AJ (1992) ALOGIT Users' Guide, Version 3.2. Hague Consulting Group, The Hague

Hensher DA, Green WH (2003) The mixed logit models: the state of practice. Transportation 30(2):133–176

Hunt JD (1990) A logit model of public transport route choice. Inst Transp Eng J 60(12):26–30

Hunt JD (1994) Evaluating Core Values Regarding Transportation and Urban Form in Calgary, BGS:22-10-94. City of Calgary GoPlan Project, Calgary, AB, Canada

Hunt JD (1996) Evaluating Core Values Regarding Transportation and Urban Form in Edmonton. City of Edmonton MasterPlan Project, The City of Edmonton, Edmonton AB, Canada

Hunt JD (2001) Stated preference analysis of sensitivities to elements of transportation and urban form. Transp Res Rec 1780:76–86

Hunt JD (2003) Modelling transportation policy impacts on mobility benefits and Kyoto-Protocol-related emissions. Built Environ 29(1):48–65

Hunt JD, Brownlee AT and Doblanko LP (1998) Design and calibration of the Edmonton Transport Analysis model. In: Preprints for the 78th Annual Transportation Research Board Conference. January 1999, Washington DC, USA

Koppelman FS (2006) Closed form discrete choice models. Handbook of transport modeling. Pergamon, Oxford, pp 211–228

McFadden D (1974) Conditional logit analysis of qualitative choice behavior. In: Zarembka P (ed) Frontiers in Econometrics. Academic Press, New York, pp 105–142

McFadden D (2007) The behavioral science of transportation. Transp Policy 14:269–274

McMillan JDP (1996) A stated preference investigation of commuters attitudes towards carpooling in Calgary and Edmonton. Dissertation, University of Calgary

Ortúzar JdeD, Willumsen LG (1994) Modelling transport, 2nd edn. Wiley, New York

Train K (2003) Discrete choice methods with simulation. Cambridge University Press, Cambridge, UK

DRAM Residential Location and Land Use Model: 40 Years of Development and Application

Stephen H. Putman

Abstract The DRAM residential model was one of the earliest to be developed and applied, with work beginning in 1971 and with applications in planning agencies continuing to this day. It was developed with the expectation that it would be applied together with an employment model (EMPAL), and with both being linked to a suite of transportation models. This chapter describes the development path of DRAM as well as those of related issues of model calibration and links to other models. The author concludes with the argument that while continued theory development is essential for models such as these, their use as forecasting and policy analysis tools depends as much upon ease of implementation for agency users as it does on any improvement in model formulation.

1 Introduction

In the U.S. the DRAM and EMPAL models of household and employment location and land use, including their successor GIS based model systems METROPILUS and TELUM, are the most widely applied models of these phenomena ever to be developed. They have seen use for public agency forecasting and policy analysis purposes in nearly 30 different metropolitan regions, including eight out of the country's ten largest cities. The development of these models began in the 1970s and continues to the present, via continued interplay between theory and practice. It would be silly to claim that these were perfect models. They do produce reliable estimates of long term regional patterns and have the ability to give sensible responses to many, though obviously not all, policy inputs. Of course other models have been developed and applied, some of them, too, began as long ago as the 1970s. In addition, it is inevitable that other operational models will be developed in

S.H. Putman
Department of City and Regional Planning, University of Pennsylvania, Philadelphia, PA, USA

F. Pagliara et al. (eds.), *Residential Location Choice*, Advances in Spatial Science,
DOI 10.1007/978-3-642-12788-5_3,

years to come. Some of these alternative approaches, both developed, and now being developed, are also presented in this book. DRAM and EMPAL, versions of which are still in active agency use more than three decades after their initial development, helped open the way.

In the following pages I provide a description of DRAM and the procedures for estimation of its parameters. The discussion begins with a description of its derivation from the principles of household location surplus maximization. The use of location surplus as an output indicator of policy effects is also described. This is followed by a discussion of the LANCON submodel's procedures for calculating land consumption by locators.

Following this is a description of CALIB, a constrained gradient search procedure, which is used to estimate the equation coefficients in both DRAM and EMPAL. This procedure calculates maximum likelihood estimates of the equation coefficients, goodness-of-fit statistics, asymptotic t-tests of the coefficients' statistical significance, and point elasticities which provide sensitivity information.

This is followed by a section on the several types of constraint procedures which are incorporated in the model structures, and then by a discussion of inter-model linkages as well as linkages to transportation models. Finally I give some concluding thoughts on the underutilization of these models in agency practice, and a new model system I have developed to address that issue.

2 The Residential Location Model: DRAM

I first began work on these models about 40 years ago. After experimenting with several functional forms, a spatial interaction structure (SI) was selected. This form was just emerging at that time as a mathematically consistent structure derived from entropy maximizing principles. In adapting the model for actual application I added the notion of a multivariate attractiveness function and a multiparametric travel cost deterrence function. This idea fell beyond the then current thinking about SI models, and required the development of new methods of model calibration.

The SI approach did, at that time, provide a nicely structured view of location models, but it lacked an appropriate grounding in economic theories of activity location. Further, the use of the multivariate attractiveness term was clearly necessary for satisfactory model application, yet did not seem to have a satisfactory basis in the model derivations. It was not until several years later that a satisfactory theoretical underpinning was developed.

It is difficult to sort out which came first, as at that time there was a great flurry of work, and many informal paper drafts were in circulation. An important paper was published by Cochrane (1975), in which a "location surplus" notion was developed. The derivation begins with the assertion that the trips which provide the trip-maker with the greatest *net* benefit are the trips that are chosen. The observed trip distribution pattern is thus indicative of the overall probability of trips being chosen on that basis. The approach taken to the subsequent derivation of a singly constrained SI model involves assuming that the probability of a particular trip-maker taking a trip from zone A to some other zone B, is the probability that a trip to zone B offers a surplus, or *net* benefit, greater than that which could be had from a trip to any other zone. It is then hypothesized that the probability that a trip to zone B for any individual trip-maker is the optimal trip (i.e., the trip which will actually be taken) increases with the number of opportunities for trip satisfaction in zone B, and decreases with trip cost between zones A and B, since the net benefit is reduced by greater trip cost.

In proceeding through the derivation of the functional form of the model it is assumed that the number of zones is large, say 100 or more. It is assumed that the underlying probability distribution is approximately exponential in the upper tail. It is shown that it is not necessary to know the actual number of trip possibilities represented by any trip attracting activity, but rather it is only necessary to assume that the number of trip possibilities is proportional to some measure of attraction.

The surplus then becomes the difference between the *probabilistic* utility u, which is the *gross* benefit of taking the trip, and a *deterministic* cost c_{ij} of actually taking the trip. Thus the *net* benefit, or surplus, from taking the trip is $s_{ij} = u_{ij} - c_{ij}$, and the probability that the surplus will be of some particular value s, given all the preceding assumptions, is

$$\Phi_{ij}(s) = exp[-h A_j \; e^{-\lambda(s-m+c_{ij})}] \tag{1}$$

where $\Phi_{ij}(s)$ is the cumulative distribution function of the "location surplus" accruing from the *optimal* trip between zone i and zone j;

h,λ,m = constants;
A_j = a measure of the attractiveness of zone j.

With this, we maintain the assumption that the trip chosen by a trip-maker will be the trip that maximizes his/her personal surplus. The probability that the trip chosen will be a trip from origin i to a particular zone j is the probability that the maximum surplus offered by a trip terminating is j is greater than the maximum surplus offered by a trip terminating in any other zone.

Continuing through the derivation, Cochrane winds up, given O_i trips originating from zone i, with the expected number of trips from zone i to zone j being given by

$$T_{ij} = \frac{O_i \; A_j \; e^{-\lambda c_{ij}}}{\sum_j A_j \; e^{-\lambda c_{ij}}} \tag{2}$$

which is the usual equation of the singly constrained SI model. Further, the total surplus for all trips actually made is

$$S_T = \frac{1}{\lambda} \sum_i O_i \, [0.577 + ln(h \, e^{\lambda m} \sum_j A_j \; e^{-\lambda c_{ij}})] \tag{3}$$

Thus beginning with rather innocuous assumptions regarding a utility maximizing basis for trip-making behaviour a robust economic underpinning can be developed for a SI model formulation. The relationship of this derivation, which is also called group-surplus maximization, to a derivation from random utility theory and probabilistic choice models is described in (Wilson et al. 1981).

The actual formulation of DRAM differs from (2) in that the attractiveness variable A_j is replaced by a multivariate formulation with a Cobb–Douglas functional form, i.e., a product form with each term being an independent variable raised to a parameter power, as in,

$$A_j = X_{1j}^{\alpha}\, X_{2j}^{\beta}\, X_{3j}^{\gamma}\, X_{4j}^{\delta} \tag{4}$$

where the X_{1j}, X_{2j}, X_{3j}, X_{4j} are attributes of zone j such as available land, household income, etc., and where α, β, γ, and δ are empirically estimated parameters.

Thus, in application DRAM is an aggregate form of a multinomial logit model of location choice. In computational form this yields the functional equivalent of a modified singly-constrained spatial interaction model. There are two modifications (1) a multivariate, multiparametric attractiveness function is used, (2) a consistent balanced constraint procedure is included in the model, allowing zone and/or locator specific constraints. The model is normally used for up to eight household categories, defined in terms of income, whose parameters are individually estimated (Putman 1983). The equation structure currently in use also makes provision for a lag term which adds stability to the model. This structure is given here:

$$N_i^n = \eta^n \sum_j Q_j^n B_j^n W_i^n c_{i,j}^{\alpha^n} \exp(\beta^n c_{i,j}) + (1.0 - \eta^n) N_{i,t-1}^T \tag{5}$$

where

$$Q_j^n = \sum_k a_{k,n} E_j^k \tag{6}$$

and

$$B_j^n = \left[\sum_i W_i^n c_{i,j}^{\alpha^n} \exp(\beta^n c_{i,j}) \right]^{-1} \tag{7}$$

and

$$W_i^n = \left(L_i^v\right)^{q^n} (x_i)^{r^n} \left(L_i^r\right)^{s^n} \prod_{n'} \left(1 + \frac{N_i^{n'}}{\sum_n N_i^n} \right)^{b_{n'}^n} \tag{8}$$

where

E_j^k = employment of type k (place of work) in zone j
N_i^n = households of type n residing in zone i
$N_{i,t-1}^T$ = total households residing in zone i at time t − 1
L_i^v = vacant developable land in zone i
x_i = 1.0 plus the percentage of developable land already developed in zone i
L_i^r = residential land in zone i
$a_{k,n}$ = regional coefficient of type n households per type k employee
$c_{i,j}$ = impedance (travel time or cost) between zones i and j
$\eta^n, \alpha^n, \beta^n, q^n, r^n, s^n, b_{n'}^n$ = empirically derived parameters

DRAM is also capable of including additional attractiveness variables in the spatial potential term, (8), of the model. There has been rather little use of this option in practice, as *the inclusion of such variables does require the subsequent development of a means for their updating* in forecast runs of the model. We have, for example, explored the merits of including residential land value as an additional attractiveness variable. We found, using rather reliable data, that the inclusion of land value in addition to household incomes made rather little contribution to the model's overall reliability, and brought with it the not insignificant prospect of having to develop an extension of the model to update the land value variable as a part of the long term forecasting procedure. In every case, when considering the addition of variables to this sort of model, one must weigh the possible improvements in model performance to be had from such additions against the cost of updating those new variables over a 30 year forecast horizon.

3 Location Surplus as an Output Measure from DRAM

Location surplus is a measure of the aggregate benefit households receive from the attributes of their chosen residential zone. Because household utility can only be measured on an ordinal scale (i.e., it is not possible to determine the monetary value of utility), the location surplus measures are interpreted as index numbers. The larger the value of location surplus, the more utility households receive from their choices of residential location.

The location surplus measures used in DRAM can be derived by using either of two different methods. Both methods produce the same location surplus measures and are based on the assumption that households attempt to maximize utility when choosing residential locations. For the first method, the DRAM model is interpreted as a multinomial logit model and the location surplus measure is found by calculating aggregate indirect utility (McFadden 1974; Ben-Akiva and Lerman 1985; Freeman 1993). In the second approach, the location surplus measure is found by directly integrating the DRAM travel demand function (Neuburger 1971; Cochrane 1975; Williams 1976).

Consider first the calculation of *location surplus from indirect utility*. In DRAM residential attractiveness is defined in (8). From this, indirect household utility is defined as:

$$V_{ij}^{n} = \ln\left[\left(c_{ij}\right)^{\alpha n} \exp\left(\beta^{n} c_{ij}\right) W_{j}^{n}\right] \tag{9}$$

where

V_{ij}^{n} = the indirect utility of a type n household that resides in zone i with a head-of-household employed in zone j,

c_{ij} = the travel time between zone i and zone j, and

α^{n}, β^{n} = empirically derived parameters.

This definition of indirect utility is for a single household. To find aggregate location surplus, it is necessary to sum the values of indirect utility for all type n households.

$$LS^{n} = \sum_{j}\sum_{k}\left(a_{kn}E_{j}^{k}\right)\ln\left[\sum_{i}\left(c_{ij}\right)^{\alpha n}\exp\left(\beta^{n} c_{ij}\right)W_{i}^{n}\right] \tag{10}$$

where

LS^{n} = the location surplus for type n households,

a_{kn} = regional coefficient of type n households per type k employee

E_{j}^{k} = employment of type k (place-of-work) in zone j.

4 Procedures for Calculation of Land Consumption: LANCON

In the combined model EMPAL and DRAM structure the use of land by locating activities is calculated after completing the calculation of total location demand. EMPAL calculates location demanded by employers, followed by DRAM's calculation of location demanded by households. LANCON takes both these calculated demands and estimates the actual change in the amount of land, by zone, that will be used by each of the demand categories. If there has been a decrease in demand by a particular demand category, then land currently in use by that category is released into a "pool" of land available for any use. If there has been an increase in demand by a particular demand category, then the addition of land to use by that category is calculated. After the calculations are done for each demand category, the sum of land used is adjusted, by an increase in densities, to match the land available for such uses.

The land used by each demand category is estimated in terms of the rate of land use by a locator in that specific demand category. The calculation, for example, of the rate of residential land use by new household locators in a specific zone is given by the following equation:

$$\frac{L_{i}^{r}}{N_{i}^{T}} = k_{0}\left(\frac{L_{i}^{d}}{L_{i}^{d}+L_{i}^{v}}\right)^{k_{1}}\left(\frac{L_{i}^{B}}{L_{i}}\right)^{k_{2}}\left(\frac{L_{i}^{C}}{L_{i}}\right)^{k_{3}}\left(\frac{N_{i}^{1}}{N_{i}^{T}}\right)^{k_{4}}\left(\frac{N_{i}^{4}}{N_{i}^{T}}\right)^{k_{5}}\left(L_{i}^{d}+L_{i}^{v}\right)^{k_{6}} \tag{11}$$

where
L_i^d = total developed land area of zone i;
L_i^v = vacant developable land in zone i;
L_i^B = "Basic" employment land in zone i;
L_i^C = "Commercial" employment land in zone i;
N_i^1 = Number of low income households in zone i;
N_i^4 = Number of high income households in zone i;
$k_0, k_1, k_2, k_3, k_4, k_5$ = empirically derived parameters.

5 Model Calibration

The calibration process involves "fitting" the equations of DRAM to the data which describe a particular region. When I first began to experiment with formulations of these models, I was unaware of the work done by others to calibrate spatial interaction models. Their efforts were devoted exclusively to making use of trip matrices (origin–destination trip data) for calibration of single parameter (usually the β in a travel function) spatial interaction model formulations. I assumed from the outset, that it would be necessary to have parameters not only in the travel function part of the model, but along with the attractiveness variables as well. Further, given the near complete unavailability of trip matrices for the cities in which I expected to do my initial model calibrations, it never would have occurred to me do use the procedures then in use by the spatial interaction modelers. Instead, I developed what we now call trip-end calibration as compared to the trip matrix based trip-interchange calibration. Beginning in about 1973 I became the scourge of various professional conferences as I pressed my colleagues to send me copies of their urban area data. Over the next 10 years I fit (statistically) the models to perhaps 40 urban areas while learning how best to do it (Putman 1977, 1980; Putman and Ducca 1978a, b). Somewhat later we were able to demonstrate that, for any specific region's dataset, the mean expected values of the parameters were identical for both trip-end and trip-interchange calibrations, though the variance, as would be expected, was somewhat higher for the trip-end procedure due to its having less information input (Putman and Kim 1984a, b).

To perform calibrations it is necessary to have one or more indicators of *Goodness-of-Fit* of the models to the data. The equation structure of DRAM is intrinsically nonlinear, and the data from which its parameters must be estimated are not normally distributed. Because of these factors it is not possible to use conventional regression techniques to calibrate DRAM. The procedure used for the estimation of parameters is gradient search. In effect, the partial derivatives of a goodness-of-fit criterion with respect to each specific parameter are calculated. The values of these derivatives determine the direction of parameter search (Putman 1983). The appropriate goodness-of-fit measure for calibration of DRAM is the likelihood function, a measure derived from the notion of maximum likelihood as developed in econometrics. This measure has the general form:

$$L = \sum_i N_i \ln \widehat{N}_i \tag{12}$$

where L is the computed likelihood measure, N_i is the observed value, and $\hat{N}i$ is the estimated value of the dependent variable. In DRAM the dependent variable would be households of a particular type located in a particular zone. It is important to note that in this equation form, the magnitude of L will be conditional on the magnitudes of the data being used. In a region with millions of households L will be larger than it will be in a region with hundreds of thousands of households. This means, unfortunately, that it is not possible to compare the results of analyses of different data sets, and thus not possible to evaluate the adequacy of one statistical analysis versus another.

The "Best Fit" is when the difference between the models' estimate of the dependent variable and the observed values in the calibration data set is as small as possible. A perfect fit would be obtained if, for each independent variable observation, i.e., locator type and zone, the estimated and the observed N_i were equal. This would give the following "Best Fit" value of likelihood:

$$L_b = \sum_i N_i \ln N_i \tag{13}$$

The "Worst Fit" would be when all values of the dependent variable are estimated by the mean of that variable. Thus, for example, if the region's total of Type 1 households were to be divided by the number of zones to get the mean of Type 1 households per zone, and all zones were assigned an amount of Type 1 households equal to the mean. This is also known as the uniform distribution assumption, where the estimated $\hat{N}i =$ the Zonal Mean $\bar{N}$, and gives the following "Worst Fit" value of likelihood:

$$L_w = \sum_i N_i \ln \bar{N} \tag{14}$$

From these two extreme values of likelihood we can, for a particular dataset, construct a relative measure of goodness-of-fit which is analogous to the more traditional R^2 measure, but which is appropriate to the nonlinear equations of DRAM and EMPAL, and to the non-normal distributions of the data. This measure of "Relative" goodness-of-fit is called a likelihood ratio, and takes the following equation form

$$\phi = \frac{L - L_w}{L_b - L_w} \tag{15}$$

The computed value of this Likelihood Ratio,ϕ, has a range such that for a perfect fit, $\phi = 1.00$, and for the worst fit, $\phi = 0.00$. Typical results obtained when fitting DRAM and EMPAL give $\phi = 0.80$–0.95. The values taken byϕ are independent of the magnitude of the dependent variables and thus it is possible to compare the calibration results of one locator type to another, or from one region to another.

6 Asymptotic t-Statistics in DRAM Calibrations

In the estimation of nonlinear model parameters it is necessary to develop ways of assessing their statistical significance. The maximum likelihood estimator, when correctly calculated, is asymptotically normally distributed with its mean equal to the true parameter value, and with a covariance matrix which can be calculated by use of second order partial derivatives. These derivatives are calculated as part of the parameter estimation procedure, and allow the computation of asymptotic t-statistics which yield an indication of the statistical significance of the individual parameters in the models' equation structures.

7 Location Elasticities for DRAM and EMPAL

Location elasticities measure the sensitivity of household location to changes in the models' attractiveness variables. All of the location elasticities are defined for a single residential zone. For a 1% increase in an attractiveness variable in a specific zone, the location elasticity measures the resulting percentage change in the number of households in that zone. For example, suppose that for low-income households in zone 12 the DRAM location elasticity for residential land is equal to 0.2500. This means that a 1% increase in residential land in zone 12 will result in 0.25% increase in the numbers of low-income households in that zone.

The location elasticities are static measures of model sensitivity. This means that when a location elasticity is calculated for a specific attractiveness variable in a specific zone it is assumed that the values of all other attractiveness variables remain fixed. In the example above, the only variable that is allowed to change is the quantity of residential land in zone 12. All of the other attractiveness variables in zone 12 are assumed to be fixed, as are the attractiveness variables (including residential land) in all zones other than zone 12. Because they are static measures of model sensitivity, the location elasticities will change as the values of the DRAM attractiveness variables change (e.g., the location elasticities for forecast years will be different from the location elasticities for the base year).

The value of the location elasticity for a specific attractiveness variable and zone is a function of (1) the value of the calibrated parameter for the attractiveness variable, (2) the numbers of households or employees in the zone, (3) the magnitude of the attractiveness variable, and (4) the relative attractiveness of other zones in the region. Location elasticities will be larger when the calibrated parameter for the attractiveness variable is large (in absolute value), the number of households or employees is small (relative to other zones in the region), or the value of the attractiveness variable is small (relative to other zones in the region). For more a more detailed description of the derivation of location elasticities for residential location models see (Anas 1982) and (Anas and Chu 1984).

Except for travel time, all of the DRAM location elasticities have the same mathematical function definition. For the percentage of developable land developed and the household percentage variables, the location elasticities are defined for changes in one plus the value of the variable. (For example, if the percentage of developable land developed equals 66%, the DRAM attractiveness variable is equal to 1.66. A 1% increase in this variable is equal to 0.0166.)

Location elasticity for any attractiveness variable (shown for residential land) is:

$$\varepsilon^{\mathrm{n}}_{\mathrm{L}^{\mathrm{r}}_{\mathrm{i}}} = \frac{\partial \mathrm{N}^{\mathrm{n}}_{\mathrm{i}}}{\partial \mathrm{L}^{\mathrm{r}}_{\mathrm{i}}} \frac{\mathrm{L}^{\mathrm{r}}_{\mathrm{i}}}{\mathrm{N}^{\mathrm{n}}_{\mathrm{i}}} = \sum_{\mathrm{j}} \left[\left(\sum_{\mathrm{k}} \mathrm{a}_{\mathrm{k,n}}\, \mathrm{E}^{\mathrm{k}}_{\mathrm{j}} \right) \left(\frac{\mathrm{s}^{\mathrm{n}}}{\mathrm{N}^{\mathrm{n}}_{\mathrm{i}}} \right) \left(\mathrm{p}^{\mathrm{n}}_{\mathrm{i,j}} \left(1 - \mathrm{p}^{\mathrm{n}}_{\mathrm{i,j}} \right) \right) \right] \tag{16}$$

where

$\varepsilon^{\mathrm{n}}_{\mathrm{L}^{\mathrm{r}}_{\mathrm{i}}}$ = elasticity of type n households to changes in residential land in zone i,

$\mathrm{a}_{\mathrm{k,n}}$ = a matrix of conversion coefficients of type n households per type k employees,

$\mathrm{E}^{\mathrm{k}}_{\mathrm{j}}$ = employment of type k (place-of-work) in zone j,

s^{n} = the calibrated DRAM parameter for residential land,

$\mathrm{L}^{\mathrm{r}}_{\mathrm{i}}$ = residential land in zone i,

$\mathrm{p}^{\mathrm{n}}_{\mathrm{i,j}}$ = the probability of a type n household, with an employed head-of-household in zone j, residing in zone i, and

$\mathrm{N}^{\mathrm{n}}_{\mathrm{i}}$ = households of type n residing in zone i.

For DRAM, the location elasticities for travel time are defined for a 1% increase in the travel time for trips from all employment zones to the specified residential zone. The equation for the location elasticity for travel time is as follows:

$$\varepsilon^{n}_{c_{j.}} = \frac{\partial N^{n}_{i}}{\partial c_{j.}} \frac{c_{j.}}{N^{n}_{i}} = \sum_{j} \left[\left(\sum_{k} a_{k,n} E^{k}_{j} \right) \left(\left(\frac{\alpha^{n}}{c_{ij}} \right) + \beta^{n} \right) \left(p^{n}_{i,j} \left(1 - p^{n}_{i,j} \right) \right) \left(\frac{c_{ij}}{N^{n}_{i}} \right) \right] \tag{17}$$

where

$\varepsilon^{\mathrm{n}}_{\mathrm{c}_{\mathrm{j.}}}$ = elasticity of type n households to changes in travel times from all employment zones to residential zone i,

c_{ij} = travel time between zones i and j, and

α^{n}, β^{n} = the calibrated DRAM parameters for travel time.

The purpose of all this is to provide a means for assessing, without the need for innumerable model runs, the relative sensitivities of locators to the different independent variables in the model structure. This knowledge, in turn, provides a means for assessing the likely degree of impact of specific policy proposals on individual locator–zone combinations.

8 The Consistent Imposition of Constraints on Location

Once having calibrated DRAM and done preliminary tests of its forecasts, it is usually necessary to add additional user knowledge to the model structure. One way in which this may be done is by the imposition of constraints on location. It is of prime importance that when constraints are imposed on particular locators in particular zones, that this does not have the effect of swamping, or overwhelming, the forecasts of location by that locator type in other zones of the region being modeled. In DRAM this precaution is taken by use of a model which is a computational hybrid of "singly constrained" zones which are not constrained, and "doubly constrained" zones which are constrained. By this means various "out of the ordinary" locations can be represented, e.g., prohibiting residential location at or too near an airport, preventing decline of a locator type as a means for representation of policy incentives which it may otherwise not be possible to represent within the model structure, or forcing forecast numbers of residents on military bases to remain at exogenously specified levels. Judiciously applied, constraints inform the model structures of "anomalies" in reality.

Four types of constraint may be applied to DRAM forecast outputs. The first type of constraint, Type I, is an absolute constraint on the number of households of a specified type in a specified zone. The second type of constraint, Type II, is an absolute constraint on the total households of all types in a specified zone. When a Type II constraint is imposed on a zone, the procedure scales the unconstrained households of each type in the zone to sum to the constraint total. If one of the household types in the zone has already been constrained to a particular value, the program attempts to maintain that value while scaling the remaining household types in the zone. The third type of constraint, Type III, is a maximum value for a particular household type in a specific zone, and only operates when the forecast of that household type in that zone exceeds the maximum. The fourth type of constraint, Type IV, is a minimum value for a particular household type in a specific zone, and only operates when the forecast of that household type in that zone is less than the minimum. Again, the program tries to avoid violating previously specified constraints. If, for example Type 1 households have been scaled with a Type I constraint in zone 14, and then are scaled as a part of a Type II constraint in zone 39, the imposition of a Type III constraint on Type 1 households in zone 73 could cause some interactions when the regional control totals were being enforced. The program constraint procedures attempt to maintain consistency throughout the various possible interactions which can arise with different combinations of constraints. Note that is possible to impose both maximum and minimum constraints on a particular locator type in a particular zone thus, in effect, providing the ability to constrain the location of that activity to fall within a predefined range. Finally, we may also impose constraints in terms of densities, by first calculating the number of households that would result in a particular density in a zone, and then applying a constraint on the zones households to limit them to the previously calculated value.

9 Linked Transportation and Land Use Model Runs

In most cases DRAM is run in conjunction with other models. That is, it is linked "behind" EMPAL, which produces forecasts of employment location, for *all* employment types, which are then used by DRAM as an input to its forecasts of household location. An important facet of the linkage between these models is the employment-to-household conversion process which provides a direct means for exogenously forecast changes in regional employment mix to produce a change in the region's income distribution. There are other socioeconomic links as well, involving regional unemployment rates, household size, and employees-per-household, which provide a means for incorporating some of these important phenomena as integral and consistent components of the forecasting process.

In addition to the linkage between EMPAL and DRAM, they are typically both used together as components of an integrated transportation and land use model system. In the initial development of these models a prototype set of transportation modeling procedures were developed for system testing purposes. In agency application other procedures for travel demand, mode split, and trip assignment are regularly used. These are typically one or another of the proprietary software packages. Various applications have combined agency developed models of travel demand and mode split, with such commercial software packages for trip assignment. While some additional work was necessary to produce seamless links from EMPAL and DRAM with these packages, my practical experience is that whenever the agency actually does wish to see the connections made, they are a relatively straight forward matter. A convex combinations procedure is added to solve the combined systems for an equilibrium solution (Putman 1991). This is an equilibrium between transportation and land use, involving both employment and residence location and land use, with travel demands (as well as mode split where the agency has the capability) and trip volumes on transportation network links, and subsequent loaded, or congested, transportation network characteristics. The procedure is straightforward and usually requires only a few iterations to converge. In practice, some agencies have found that a partial equilibrium between transportation and land use will suffice to improve forecasting reliability. The partial equilibrium is measured in terms of mean absolute percentage change in activity locations or network travel characteristics from one system iteration to the next. These days computation time is often of no importance in either case, but still, it has been possible to achieve quite good results with, say three iterations rather than the four or five that might be run to reach equilibrium. It is worth noting for the record that research by Shen (1995) has proven the existence and uniqueness of the solutions obtained (a reassuring result, even though in practice it had previously been demonstrated that the solutions were computationally stable and unique).

10 Concluding Thoughts

The development of DRAM, and its companion models was originally undertaken with the intention of bringing the best practical technology into regular agency use. It has taken a great deal longer than I expected for this to happen. Today, even though there have been a substantial number of agency users, and even though more seem to become interested with each passing year, the use of statistically valid models for producing the land use inputs to travel modeling still is not regular practice amongst Metropolitan Planning Organizations (MPOs) in the U.S. In part this is a political matter. Land developers are often closely connected to politicians. Thus, while transportation modeling is used by virtually all MPOs, the statistically valid modeling of land use, even though required as input to the transportation models, is often pushed aside by political considerations. This, however, is not the sole reason for the underutilization of these methods. Other reasons derive from both theory and practice.

First, no matter how good or bad the theoretical underpinnings, what is important in practice is whether a model system is implementable and understandable. It helps if the underlying theory, albeit often incomplete, is comprehensible to the users. Often, however, model users in operating agencies are rather less concerned with the model's theoretical basis than they are with whether or not it can readily be calibrated and its outputs adjusted where necessary to meet agencies' political constraints. Only after a substantial user community evolves, do the model developers find it possible to address questions of model improvement. Even then, the disparities between different agencies' priorities for model developments and improvements can sometimes be insurmountable with limited budgets. The size of the user community in transportation modeling is probably two orders of magnitude greater than the size of the land use modeling user community. This makes a big difference in what can and should be done, and the ways in which it may happen.

Second, one of the apparent determinants of the success of a model application is the extent to which the work is being done by agency staff or by consultants. Rarely do agency staff have the training necessary to do this work, yet if agency staff are not intimately involved, there will be issues of credibility of the results within the agency. At the same time, agency staff who may be inexperienced in the use of these sorts of models have a difficult adjustment to make between spreadsheet programs and complex models of socioeconomic phenomena. The fact that it is not possible to, in effect, simply press a button and get a valid result means that new users will always have an initial period of frustration. Some simply quit at that point, with projects going unfinished and the blame being placed on the difficulty of use and/or inadequate reliability of results. Not too many years ago I was involved in a project for a city in Florida. The entire effort was plagued by serious data availability and reliability issues. It seemed that after considerable struggling, both the agency staff and their politically motivated external advisory committee had understood the data issues and the effects on forecasting and policy analysis which data problems might be expected to have. Then, at the very first public presentation which I saw of the modeling project results, the opening speaker (an agency staff member worried about intra-agency political conflicts) walked to the podium and began his presentation with the sentence: “It doesn’t work.”

After more than 30 years’ experience of straddling the model development – model application fence, I became convinced that major improvement in planning practice was not to be had from making the land use models better. After all, existing models such as DRAM are often able to account for 90% or more of the variation in household location patterns. Instead of trying to add a few percent more to the explained variation, it seemed that it would be more effective to try to decrease the vast number (the majority) of agencies that used no statistically sound models at all. This led to the development of a new model system, called METROPILUS, (Putman and Chan 2001) which contains DRAM and EMPAL, and other programs including calibration procedures, embedded in a GIS environment, operating behind an extensive graphical user interface (GUI). The principal aim in its development was to achieve a dramatic reduction in the difficulty ordinarily encountered in land use model application. This system was quite successful in allowing agency users to do their own model runs. The system was usually set up with consultant assistance. Preliminary calibrations were sometimes also done with consultant assistance. After the initial system had been set up and calibrated they were able to do this with minimal additional assistance from consultants. METROPILUS, incorporating DRAM, continues in use by several regional agencies to this time.

Even so, it still was not “user friendly” enough for some users. The world had become used to general spreadsheet and database programs, as well as more specialized applications such as income tax preparation programs, touch screen computers in convenience stores for ordering sandwiches, and to restaurant waiters or waitresses placing patron’s meal orders with touch screens. Thus we developed an even more user friendly system, named TELUM. With this, after the user enters a few preliminary numbers, a spreadsheet, dimensioned to the user’s region, is opened. The task of data collection is presented as a structured set of steps, with software assistance, towards populating this spreadsheet with the region’s data. Various consistency checks are performed automatically during the user’s progress towards assembling the required data. The program also checks the consistency of the links between the database and the geographical data by which it is embedded in the GIS. On the completion of the data assembly work, statistical analyses of the data are performed and evaluated by the model system, followed by the running of a fully automatic procedure which performs the calibration of the models. After this, the model software rearranges the data along with the statistical analyses, and notifies the user of the results, while at the same time completing the preparations necessary for making forecasts using the models. Then, literally at the press of a button, the system runs the forecasts for the user’s region.

We designed and constructed this new model system around a knowledge based systems approach (Pozoukidou 2005). Implementing this kind of automation for a complex land use modelling system had not been done before, and we had some work to do to understand the function and use of software wizards and other artificial intelligence notions. As such, the early version of this model system had several simplifications over the full form of METROPILUS, including some limits on numbers of analysis zones and numbers of locator categories. Also, certain optional procedures available in METROPILUS were temporarily disabled in TELUM. Even so, the new system, now operational, does a remarkable job of assisting agency staff, often inexperienced in location modelling work, in performing this rather complex planning/ analysis activity (Pozoukidou 2006, 2007).

We developed TELUM under contract to the New Jersey Institute of Technology (NJIT), who subsequently have overseen its deployment. Development of the TELUM model system, incorporating DRAM, was sponsored by the USDoT, and TELUM has since been distributed to every MPO in the US. The software and documentation are available from an NJIT website (NJIT 2009). In pre-release tests by regional agencies participating in "beta testing," staff from MPOs for several mid-sized regions were able to produce statistically reliable, replicable, forecasts for their agencies' use, without the need for consulting assistance. Since then several more agencies have downloaded the software and produced their own forecasts (Casper et al. 2009). In order to achieve this level of automation a modest reduction in the models' flexibility of use has been necessary, but otherwise these agencies have done, largely on their own, work that has, hitherto, always required outside consulting assistance and a major budgetary commitment. All told, a half dozen or so agencies have downloaded and applied the TELUM system in areas such as Colorado Springs, Little Rock, and Des Moines. The results have been mixed, some quite satisfactory to their users, some not so. A major issue seems to be the availability of appropriate data, and the agency staff's ability to manage data issues as they arise. Even so, agencies continue to show interest in at least attempting to use the model for their specific local forecasting and policy analysis needs. We shall see whether having thus greatly reduced the difficulty of use, land use models will be more commonly implemented as a part of planning agencies' forecasting and policy evaluation procedures. This question will be settled not, as some are wont to say, on whether one model provides a few percent better goodness-of-fit than does another, but more on a mix of internal agency politics, regional politics, and the extent to which the whole affair is run by the region's moneyed interests.

References

Anas A (1982) Residential location markets and urban transportation: economic theory, econometrics and public policy analysis. Academic Press, New York

Anas A, Chu C (1984) Discrete choice models and the housing price and travel to work elasticities of location demand. J Urban Econ 15:107–123

Ben-Akiva M, Lerman SR (1985) Discrete choice analysis: theory and application to travel demand. MIT, Cambridge
Casper CT et al (2009) Application of TELUM at Pikes Peak Area Council of Government Metropolitan Planning Organization: Lessons Learned, Paper 09-2697, presented at Transportation Research Board Annual Meeting, Washington, DC. http://www.trb.org/am/ip
Cochrane RA (1975) A possible economic basis for the gravity model. J Transp Econ Policy 9:34–49
Freeman AM (1993) The measurement of environmental and resource values: theory and method. Resources for the Future, Washington DC
McFadden D (1974) Conditional logit analysis of qualitative choice behavior. In: Zarembka P (ed) Frontiers in econometrics. Academic Press, New York, pp 105–142
Neuburger H (1971) User benefit in the evaluation of transport and land use plans. J Transp Econ Policy 5:52–75
NJIT (2009) http://www.telus-national.org/documentationpaper/htm
Pozoukidou G (2005) Increased usability of urban and land use models. The role of knowledge based systems in facilitating land use forecasting to planning agencies, Ph.D. dissertation viewable at. http://repository.upenn.edu/dissertations/AAI3179791
Pozoukidou G (2006) Planning support systems' application bottlenecks, European Regional Science Association, ERSA conference papers, No. ersa06p769 http://www-sre.wu-wien.ac.at/ersa/ersaconfs/ersa06/papers/769.pdf
Pozoukidou G (2007) Facilitating land use forecasting in planning agencies. In: Kungolos A, Brebbia CA, Beriatos E (eds) Sustainable development III, vol 1. WIT, Southampton, pp 67–80
Putman S (1977) Calibrating a disaggregated residential model – DRAM. In: Batey P, Massey D (eds) Alternative frameworks for analysis, vol 7. Pion, London, pp 108–124
Putman S (1980) Calibrating urban residential location models 3: empirical results for non-U.S. cities. Environ Plann A 12:813–827
Putman S (1983) Integrated urban models: policy analysis of transportation and land use. Pion, London
Putman S (1991) Integrated urban models 2: new research and applications of optimization and dynamics. Pion, London
Putman S, Chan S-L (2001) The METROPILUS planning support system: urban models and GIS. In: Brail R, Klosterman R (eds) Planning support systems: integrating geographic information systems, models, and visualization tools. ESRI, Redlands, California, pp 99–128
Putman S, Ducca FW (1978a) Calibrating urban residential models 1: procedures and strategies. Environ Plann A 10:633–650
Putman S, Ducca FW (1978b) Calibrating urban residential models 2: empirical results. Environ Plann A 10:1001–1014
Putman S, Kim Y-S (1984a) Calibrating urban residential location models 4: effects of log-collinearity on model calibration and formulation. Environ Plann A 16:95–106
Putman S, Kim Y-S (1984b) Calibrating urban residential location models 5: a comparison of trip end and trip interchange calibration methods. Environ Plann A 16:1649–1664
Shen P-N (1995) Optimized network equilibrium models of combined travel and residential location choices. Ph.D. Dissertation, Department of City and Regional Planning, University of Pennsylvania, Philadelphia
Williams HCWL (1976) Travel demand models, duality relations and user benefit analysis. J Reg Sci 16:147–166
Wilson AG et al (1981) Optimization in Location and Transport Analysis. Wiley, Chichester, Sussex

The DELTA Residential Location Model

David Simmonds

Abstract This paper describes the residential location component of the DELTA package. The introduction puts this in context, by very briefly describing the objectives and scope of the package, and the set of sub-models which it contains. The second section presents the residential location model itself. Subsequent sections discuss the calibration of the model, its applications, and current developments.

1 Objectives and Scope of the DELTA Package

The DELTA land-use/economic modelling package has been developed by David Simmonds Consultancy since the mid-1990s. The main objectives in creating a new package were to create a practical tool to forecast urban and regional change, and in particular to examine the expected impact of transport change; to provide a land-use/economic model which works in interaction with any appropriate transport model, and can therefore be used to extend relatively conventional transport models into land-use/transport interaction; and in doing this, to draw not only upon previous modelling experience but also upon the wide range of relevant research carried out in geography, urban economics, etc.

A full description of the thinking behind the original version of the model can be found in Simmonds (1999). The model as outlined in that paper is essentially the 1995–1996 prototype. For more recent descriptions, see Simmonds and Skinner (2004) or Simmonds and Feldman (2005).

D. Simmonds
David Simmonds Consultancy, 14 Millers Yard, CB2 1RQ Cambridge, UK
e-mail: david.simmonds@davidsimmonds.com

F. Pagliara et al. (eds.), *Residential Location Choice*, Advances in Spatial Science,
DOI 10.1007/978-3-642-12788-5_4,

2 The DELTA Sub-models

Because land-uses take a long time to respond to transport changes, the land-use model needs to represent change over time, in contrast with the conventional approach in transport modelling which describes transport supply and demand as in equilibrium at one point in time. Another requirement is to recognize that different processes operate at the urban and at the regional levels: for example, different factors affect the total economic activity in an area and the location of employment within it. The model accordingly contains urban processes which represent changes in or between the individual zones, and regional processes in which the units are whole travel-to-work areas.

The urban processes represent both changes in buildings and changes in activities. The processes of physical change are development: the amount of floorspace by zone and type (residential, retail, office, industrial, plus more specialised types in some applications). Development is driven by the economic scenario and the modelled property market, and controlled by inputs measuring what is allowed by the planning system; and the housing quality model, which models the way in which an area may decline to slum status, or be revived from slum to high-quality.

The urban activity sub-models are: the transition model, which represents household (and hence population) changes in terms of movements through a simplified lifecycle; the car-ownership model, which predicts the changing proportions of households by type and zone owning 0, 1 or 2+ cars, mainly in response to increasing incomes; the location model, which locates or relocates a proportion of households and of employment in each year, and also models the property market within which location occurs; and the employment status model, which updates the work status of residents and the commuting pattern in response to the spatial changes in households and employment.

At the regional level, there are three models, all of activities: the migration model moves households between areas; the investment model allocates investment to areas (taking account of changes in accessibility and property costs); and the production/trade model (a spatial input–output model) estimates production by sector and area and the patterns of trade between areas. All the modelled processes are considered in a fixed sequence within each one year step, as shown in Fig. 1. However, there are also numerous time lags between the different processes which are equally or more important to the overall performance of the model. As will be seen below, the residential location sub-model is affected, with time lags, directly by all the other urban level components, and indirectly by the regional level components.

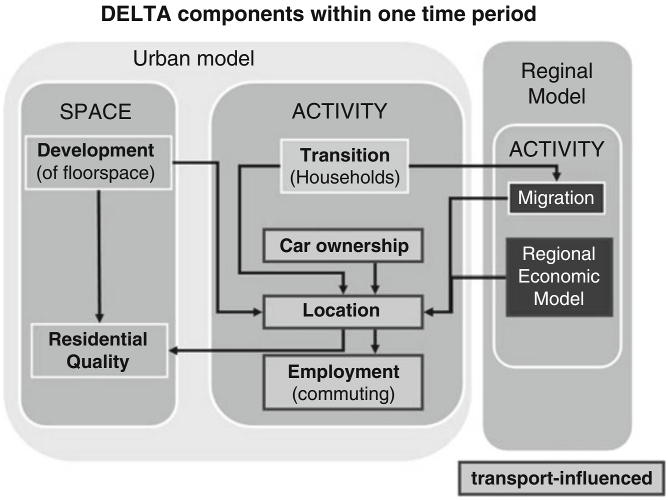


Fig. 1 The DELTA sub-model in one year

3 Specification of the Residential Location Model

3.1 Scope of the Residential Location Sub-model

The location sub-model is both the "location and relocation sub-model", and the "property market sub-model". Mobile activities respond to changes in five variables: accessibility, quality of the local environment in general, quantity of housing; quality of housing; and the cost or utility of consumption, i.e. of spending income on housing, travel, and other goods and services.

The accessibility variables are calculated from previous land-uses and from outputs of the transport model. The environmental variable is in practice also based on transport model outputs, though it could be more broadly specified. The quantity of housing is determined partly by vacancy, partly by the completion of new stock (the development model) and partly by the number of existing occupiers seeking to move.

Area quality is adjusted by the quality sub-model (see below). Utility of consumption is calculated within this sub-model. Utility of consumption is influenced by the rent of floorspace in each zone, and therefore has to be recalculated within each step of the sub-model as rents are iteratively adjusted. The rent adjustment seeks to equate the total demand for each type of floorspace in each zone (determined by the number of locating activities and the amount of floorspace each occupies, both of which are variable) with the amount of floorspace available (the total stock less that left vacant or occupied by "immobile" activities). DELTA allows for a variable proportion of the stock to remain vacant. The market mechanism uses a consumption function and rent-adjustment process similar to those in

Table 1 DELTA applications including residential location

Core area/study	Transport modelling package used	Comments	Reference(s)
Edinburgh and Lothian (1)/DELTA development, EPSRC research	START	Urban level model only	Simmonds and Still (1999)
Greater Manchester strategy planning model (GMSPM)/ various	START	Urban level model only	Copley et al. (2000), Whitehead et al. (2006)
Trans-Pennine corridor/ Strategic environmental assessment in the Trans-Pennine Corridor	START	Urban/regional model	Simmonds and Skinner (2001)
South & West Yorkshire (SWYSM)/ SWYMMS, Eddington Study and other projects	START	Urban/regional model	Simmonds and Skinner (2002, 2004), Feldman et al. (2007)
Edinburgh & Lothian (2)/ PROSPECTS	START	Urban level model only	Minken et al. (2003)
Glasgow & Clyde Valley/ CSTCS, Clyde Corridor Study, appraisals of major motorway schemes	TRIPS (CSTM 3A application)	Urban/regional model; regional model covers whole of Scotland	
Edinburgh & Lothian (3)/ New Transport Initiative	TRAM	Urban/regional model; regional model covers whole of Scotland	Simmonds et al. (2005)
Strathclyde(SITLUM)/ for use by transport and land-use planning agencies	STM	Urban/regional model; regional model covers whole of Scotland	Aramu et al. (2006)
TELMoS (Transport/Economic/ Land-use Model of Scotland)/ Lowland Scotland	TRIPS	Urban/regional model; regional model covers whole of Scotland	Nicoll et al. (2006)
Auckland (New Zealand)	EMME/2	Urban level model only	
Thames Gateway South Essex	STM	Urban model with partial regional model	Dobson et al. (2007)

MEPLAN and TRANUS (see Hunt and Simmonds 1993). The critical differences are that in DELTA they affect only the households and housing that are "in the market" during this period and they work in an incremental form rather than in a cross-sectional form. The early applications of DELTA worked with a simple set of zones in a closed Study Area. From the Trans-Pennine model onwards (see Table 1), the model works on a number of labour/housing market areas. Households move between areas primarily through the longer-distance migration process.

3.2 *The Households To Be Located*

DELTA is intended to be applied with a detailed classification of households reflecting household composition, age of household members, working status of working-age adults in the household, and the socio-economic group to which the household belongs. An important characteristic of the model is that only a proportion of households make residential choices in any one period. It is assumed that the main reasons for making a new residential choice are linked to change in one of the household classification variables, e.g. a change in the household's composition or in its work status. All households which are forecast to move from one category to another – plus a proportion of unchanged households – are assumed to make such a choice, and hence enter into the location process.

The households in the location model fall into two groups: "pool" households, which have no previous location within the area, and "mobile" households, which do have a previous location within the area modelled. Newly formed households and households resulting from existing households merging (e.g. singles forming couples) are assumed to make new location decisions and are counted as "pool" households. "Mobile" households are those which are undergoing other changes (mainly from couple with children onwards). In addition, a proportion of non-changing households is assumed to be "mobile" in each period. The numbers of "mobile" and "pool" households are initially calculated in the household transition model (which also finds and subtracts the numbers of households which have dissolved or migrated out of the modelled area altogether). The inter-area migration model is then applied, before the location model. The migration model predicts moves of households between areas within the modelled system: these households are subtracted from the "mobile" and "pool" numbers for the areas they leave, and added to the "pool" numbers for the areas into which they migrate. Households migrating from the rest of the world are also added to the "pool" numbers. The task for the residential location model in each area is therefore to locate:

$H(P)_{pa}{}^{h}$	the total "pool" of households type h to be located in area a

and to relocate:

$H(M)_{pi}{}^{h}$	the number of mobile households type h initially located in zone i.

3.3 Specification of the Model: Location Equations

The main location equations are weighted incremental logit functions, with slightly different forms for "pool" and for "mobile" households. The equation to locate "pool" households in area a to zones i is:

$$H(LP)_{pi}^{h} = H(P)_{pa}^{h} \cdot \frac{H_{ti}^{h} \cdot (F(V)_{pi}^{H} / F_{ti}^{H}).exp(\Delta V_{pi}^{h})}{\sum_{i \in a} H_{ti}^{h} \cdot (F(V)_{pi}^{H} / F_{ti}^{H}).exp(\Delta V_{pi}^{h})},$$

where:

$H(LP)_{pi}{}^{h}$	households type h located from the "pool" of locators to zone i during period p,
$H_{ti}{}^{h}$	total households type h living in i at time t (the beginning of the modelled period),
$F(V)_{pi}{}^{H}$	total available housing floorspace in i during period p,
$F_{ti}{}^{H}$	previous occupied housing floorspace in i and
$\Delta V_{pi}{}^{h}$	change in the utility of locating in zone i for households of type h locating during period p.

The hypothesis here is that if there are no changes in floorspace or utility of location in any zone, the newly-locating households of type h will distribute themselves in proportion to similar households already located. The model therefore assumes that there is a set of unmodelled variables at work, including for instance details of housing types and area characteristics, which will tend to draw households of particular types to particular zones. The equivalent equation for "mobile" households is:

$$H(LM)_{pi}^{h} = \left\{ \sum_i H(M)_{pi}^{h} \right\} \frac{H(M)_{pi}^{h} . (F(V)_{pi}^{H} / F(M)_{ti}^{H}) . exp(\Delta V_{pi}^{h})}{\sum_i H(M)_{pi}^{h} . (F(V)_{pi}^{H} / F(M)_{ti}^{H}) . exp(\Delta V_{pi}^{h})},$$

where the additional variables are:

$H(LM)_{pi}{}^{h}$	mobile households type h located to zone i,
$F(M)_{ti}{}^{H}$	initial "mobile" housing floorspace, i.e. the space previously occupied by all the households now classified as "mobile".

This equation implements the subtly different hypothesis that "mobile" households will tend to remain where they are unless the floorspace or utility of location variables change and modify their preferences. This means, for example, that if in one year a zone attracts a significant number of households of the more stable types, e.g. households with children, the subsequent history of the zone will be influenced by the tendency for those households to stay there and eventually to change (children grow up and leave, adults retire) in situ. Note that the model does not distinguish between "staying in the same dwelling" and "moving to another dwelling within the same zone". The following sections explain the variables affecting location within the above equations.

3.4 Floorspace Variables

The total available housing floorspace in each zone is equal to the current total stock of built floorspace (inclusive of newly completed developments) *minus* that occupied by immobile households, i.e. those which are not considering a move in this time period; floorspace which is being held vacant by landlords; and any other floorspace which is being vacated for policy purposes (usually as a preliminary to demolition and redevelopment). The amount of floorspace held vacant by landlords depends on the rents being offered, and is adjusted as part of the iterative process within the residential location model for each year (see *Solution process* below). The previous occupied housing floorspace is the quantity of housing occupied by households in the previous year, i.e. the total stock in the zone at that time less any which was vacant. The oddly-named category of "mobile" housing floorspace is the quantity of floorspace initially occupied by households which have been classified as "mobile"– see above.

3.5 Location Sub-model: Households' Change in Utility of Location

The term "change in utility of location" summarises all the explicitly modelled factors affecting the locational preferences of each household type apart from the physical quantity of housing floorspace. This term is calculated as:

$$\Delta V_{pi}^{h} = \theta_{p}^{Uh} \,.\, \Delta U_{\Delta t,i}^{h} + \theta_{p}^{Ah} \,.\, \Delta A_{\Delta t,i}^{h} + \theta_{p}^{Qh} \,.\, \Delta Q_{\Delta t,i}^{h} + \theta_{p}^{Rh} \,.\, \Delta R_{\Delta t,i}^{h},$$

where:

$\Delta U_{\Delta t,i}{}^{h}$	change, over a defined past period Δt, in utility of consumption for households type h locating in zone i,
$\Delta A_{\Delta t,i}{}^{h}$	similar past change in accessibility of zone i for households type h,
$\Delta Q_{\Delta t,i}{}^{h}$	similar past change in quality of housing areas in zone i,
$\Delta R_{\Delta t,i}{}^{h}$	similar past change in transport-related environmental quality as perceived by households type h in zone i,
$\theta_p{}^{Uh}$, $\theta_p{}^{Ah}$, $\theta_p{}^{Qh}$, $\theta_p{}^{Rh}$,	coefficients on the above.

The utility of consumption term (note the difference between utility of consumption and utility of location) represents the utility which households obtain from spending their incomes given their choice of location. This term therefore depends on households' spending preferences, and is to some extent determined by the household itself. The other three terms – accessibility, environmental quality and environmental quality – are more in the nature of externalities, i.e. individual households cannot directly purchase better or worse values of these but must choose and typically make trade-offs between the levels on offer in different zones. The following sections explain these four variables in turn. All four variables within the utility of location term are measured over a past time period Δt, which is different for different household types. The length of this period is based on the average time that households of type h have lived in their present dwelling. This value is very low for young single persons, and much higher for older, more settled households. The effect of these definitions on response to accessibility change is that impacts of an "external" change (e.g. in the transport system) on highly mobile households, such as those of young single persons, will be felt very rapidly; impacts on less mobile households will be spread over a number of years. Once the initial, direct impact of the change in accessibility has been felt, the effect tends to persist through the incremental formulation of the location model as described above. For example, if the effect of a transport change is to encourage more young single people (perhaps of a particular socio-economic group) to live in a particular zone, then subsequent "cohorts" of similar people will (other things being equal) be correspondingly more likely to choose to live in that zone.

3.6 Utility of Consumption Variables

DELTA's treatment of the relationship between households and housing closely follows the "Martin Centre tradition" (see Simmonds 1994) in assuming that housing is a continuous variable of which households can choose to occupy larger or smaller quantities within any one zone. This flexibility subsumes both the choice of dwelling types within each zone, and the scope for households to adapt housing to their needs by arrangements such as sub-division, multiple occupation or sub-letting. DELTA also follows the Martin Centre tradition (and many other urban models) in assuming that all floorspace is rented rather than purchased. The quantity of floorspace which a household will choose to occupy, conditional on choosing to live in zone i, is calculated by maximising a Cobb-Douglas type utility function:

$$U^h_{(t+1)i} = (a^{hH}_{pi})^{\alpha^{hH}_p} \cdot (a^{hO}_{pi})^{\alpha^{hO}_p},$$

where:

${a_{pi}}^{hH}$	space per household type h in zone i (see calculation below),
${a_{pi}}^{hO}$	expenditure on other goods and services per household type h in i (see calculation below),
${\alpha_p}^{hH}$, ${\alpha_p}^{hO}$	propensity of households type h to spend available income on housing, H, or other goods and services, O.

Note that the model requires that ${\alpha_p}^{hS} + {\alpha_p}^{hO} = 1$ for each household type h, i.e. that all of the modelled income must be spent either on space or on other goods and services. The model finds the maximum utility of consumption for each type of household in each zone, given the income of that household type and the current rent of housing in the zone. Adjustments are made to allow first for the fact that certain household types enjoy housing subsidies and therefore occupy more space than they can apparently afford, and secondly for other unmodelled variations in the base year ratio of households to housing.

3.7 Accessibility Variables

The accessibility variables used in the residential location model are weighted sums of a range of more specific accessibility measures. The specific accessibility measures include accessibility to work opportunities, to shop opportunities and (in most applications) a range of other measures. Accessibility to work is (again in most applications) disaggregated by socio-economic group, so that households classified as "unskilled manual" are influenced only by accessibility to "unskilled manual" jobs. Each individual accessibility measure is of the form:

$$A_{ti}^{po} = \frac{1}{-\lambda_t^{Dp}} \left(\ln \left\{ \sum_j W_{tj}^p \exp\left(-\lambda_t^{Dp} g_{tij}^{po} \right) \right\} - K^p \right),$$

where:

λ_t^{Dp}	is the destination choice coefficient for purpose p at time t,
W_{tj}^p	is the measure of opportunities for purpose p in zone j at time t,
g_{tij}^{po}	the generalised cost of a tour from i to j and return, for purpose p and car-ownership level o at time t,
K^p	a constant for purpose p (defined as the logarithm of the base year sum of W_j^p).

The weighting of accessibilities reflects first of all the typical trip-making frequency of the household type in question. There is one exception in this weighting: households with unemployed members are assumed to have weights based on the numbers of trips they would make if those unemployed members were actually in work – i.e. households which are seeking work are influenced by accessibility to job opportunities in the same way as those which have work. (Their eventual location of choice will of course be different, because households with fewer or no employed members have lower budgets to support these preferences.)

The second stage of weighting is simply based on the car-ownership of households of type h in zone i at each point in time. This treatment of car-ownership in accessibility is consistent with the assumption that location and car-ownership are joint choices (though car-ownership can also change for non-relocating households.) For the model to behave reasonably, it is essential that the measures of accessibility should reflect the greater accessibility which (except in very unusual circumstances) will result from greater car-ownership. It is also essential that they should include walking as a viable mode of transport for short or very short journeys – without a "walk" option, the advantage of locating close to attractor land-uses will be understated.

The accessibility terms thus bring into the location model the effects of change in transport levels of service, changes in the location of the other land-uses with which households need to interact, and changes in household car-ownership. The accessibility terms are therefore critical to the impacts of transport on land-use.

3.8 Environmental Variables

The DELTA package is designed to be capable of interfacing with an environmental model which would ideally take account of emissions both from transport and non-transport sources, model the dispersion of pollutants (and all the factors affecting this) and hence calculate the immissions affecting residents in each location. In practice, the opportunity to link to such an environmental model has not yet arisen, and the environmental variables used within DELTA are based directly on the transport model outputs, without reference to other sources of pollution or to dispersion. (More sophisticated environmental analysis has been carried out in some of the studies using, as input, the outputs from DELTA and from the associated transport model; see for example Coombe et al. (2001). However, that analysis is not [yet] fed back into DELTA and is therefore not relevant here.) Two different approaches to transport–environmental variables have been used: using zonal values for a range of different indicators, with appropriate weights in each indicator, or using a single traffic-density measure (passenger-car-unit-Km per Km^2) as a proxy for the range of traffic impacts. The explicit indicators used have typically been noise, carbon monoxide, oxides of nitrogen and volatile organic compounds. The advantage of this approach is that it allows different indicators to be given different weights by household type – for instance, some households may be more sensitive to noise changes than others. There are however problems in calculating meaningful zonal values for some of these; and for others there is the problem that it is not necessarily reasonable to assume that people can respond directly to the level of concentration of an odourless, colourless gas, however significant it may be in health terms. We have therefore avoided these issues in some studies by using the traffic-density measure; this excludes modelling the likelihood that some of the adverse impacts of road traffic will be reduced by future changes in vehicle technology, but rightly puts more emphasis on the risk and severance effects of traffic itself. There are still limitations to the appropriateness of zonal values, but it is reasonable to assume that risk and severance effects apply throughout residential areas, not just to the residents of dwellings fronting major roads.

3.9 Quality Variables

The housing quality variable is an index. This index is adjusted in the quality model to reflect the impact of households and occupation/vacancy on the housing stock itself. The standard operation of the quality model is that the index tends to rise if housing in a zone is fully occupied by high-income households, and tends to fall if it is occupied by low-income households or (worse) left vacant. This is a potentially important feedback effect, the more so as it can be a positive feedback: if, say, a transport change has the effect of increasing the proportion of higher-income households in a lower-quality zone, their presence (and their expenditure) will gradually improve the quality of the zone and attract yet more higher-income households.

3.10 Solution Process

The inner operation of the location model in respect of households and housing is to calculate utility of consumption for each household type in each zone, given the current rent per unit housing floorspace (in the first iteration, the previous period's rent); in doing so, find the amount of space that each household will occupy if it locates in each zone; calculate the change in utility of location for each household type in each zone; locate all locatable households; find the total space used by these households, compare with the available space, and if necessary adjust the rent and repeat. An additional outer loop adjusts the quantity of housing which is being held vacant, in response to the changes in rent. This modifies the available floorspace, and further iterations of the inner loop are then carried out to solve the model with the modified supply.

4 Calibration and Validation of the Residential Location Model

4.1 Approach

The approach to the residential location model (as to most of DELTA) has been very much one of designing a sensible model structure and attempting to complete it (with detailed variable definitions, with some minor variation of functional forms and with coefficients) in the light of a variety of previous research, supplemented when and where possible by new calibration. This contrasts with the alternative approach in which (in the extreme) the model design is strongly influenced and the coefficients are wholly determined by the results of calibration using data for the study area in which the model is being applied. More formal calibration of DELTA has been considered, but opportunities to pursue this to any significant degree are only now being found. Likewise, formal validation of DELTA's performance over time would be desirable, but no opportunity to carry it out has yet been realised. This section therefore concentrates on identifying the previous research which has been considered in finalising the various DELTA residential location applications.

4.2 Sources Used

The original Edinburgh prototype was implemented starting with coefficients on *utility of consumption* and *accessibility* estimated as part of an earlier cross-sectional calibration on data for Bristol. These coefficients were estimated at a relatively aggregate level, and a process of interpolation and extrapolation was needed to obtain values for each of the household types in the DELTA model. (It would of course have been possible simply to apply the more aggregate coefficients to all the corresponding household types, but that was considered unnecessarily crude.) That process provided the two coefficients needed for the residential location model to work at all within the DELTA framework, i.e. for households to respond both to changes in accessibility (and hence to transport – necessary for land-use/transport interaction) and (through the utility of consumption term) to changes in housing rent (necessary for convergence of the location model itself). Given the relationship between utility of consumption and households' expenditure on housing, it is possible to derive coefficients for any other variables from exogenously researched willingness-to-pay values for those variables. This approach has been used to obtain the coefficients for *housing quality* and for *environmental quality*. Both of these variables are defined so that a unit increase in the variable produces a 1% increase in rent in an average zone. In the case of the housing quality variable, the housing quality index is defined so as to correspond to this specification. In the case of the environmental quality variable, the weights on the different components of the variable (noise, different contaminants, etc) were defined so as to reproduce willingness-to-pay values.

In the original Edinburgh prototype, the weight on noise was set so that a localised 1 dBA increase in noise would produce, on average, a given decrease in rents, on the basis of relationships reported in Tinch (1995). The weights on the different components of air quality were calculated using two pieces of information. Firstly the relative toxicity of different emissions was considered as a means of estimating their relative importance; and secondly, willingness-to-pay for reduction in "atmospheric pollution" was used to scale the composite environmental variable. Since the model itself works in terms of utility of consumption, rather than income itself, the exact transformation is to ensure that a reduction in pollution for which a household is willing to pay x brings about the same increase in utility of consumption as an increase in income of x.

Much of the information used in determining the environmental coefficients in the Edinburgh prototype was drawn from Tinch (1995). All of the coefficients were subsequently revised in the light of results from a DELTA-related Stated Preference study carried out by the University of Leeds Institute for Transport Studies (see Wardman et al. 1997). A certain amount of adjustment has been carried out in subsequent projects in the light of other information; we are currently considering, for example, the results obtained by Pagliara and Preston (2003), and hope to make use of results from research being carried out at Napier University (Edinburgh).

5 Application of the Residential Location Model

5.1 Completed Applications

DELTA residential location models have been implemented within land-use/transport interaction models for Edinburgh (separate projects in 1995/96 and 2001/02); Greater

Manchester; the Trans-Pennine Corridor; South and West Yorkshire; the Central Scotland Transport Corridors Study (CSTCS) area (the Glasgow conurbation and Lanarkshire); Strathclyde (a separate project again focussed on West Central Scotland); the whole of lowland Scotland (the Transport/Economic/Land-use Model of Scotland (TELMoS), replacing the CSTCS model). Table 1 shows the transport models used in each of these, and references to papers or reports describing the projects in general. DELTA residential location models have also been implemented as components of land-use-only models are (without the feedback from transport) in Harlow (Essex), and Derby; and new LUTI models are being implemented for South Essex (Tilbury–Basildon–Southend), Auckland (New Zealand), London and Leicester.

6 Example Results

6.1 Introduction

It would be possible to extract results that showed the operation of the residential location model in isolation, either from the intermediate outputs within a single forecasting year or by "switching off" all the other components (as far as this is logically possible). However, this would not seem particularly informative, and we have therefore extracted some residential location results from a full DELTA application.

The results below are taken from initial runs of the SITLUM application, and were carried out purely to demonstrate the performance of the DELTA model and not to test real schemes or policy proposals. The results are discussed in terms of illustrating the model behaviour rather than of policy appraisal.

6.2 Housing Policy Impacts

The results considered here are the impacts of a test in which a substantial additional supply of housing was permitted in a corridor running eastwards from the edge of Glasgow City Centre to the periphery of the conurbation. All of the results are considered in terms of the differences between this test and a Reference Case without the additional permissions for housing development. This additional supply was actually introduced into five zones; we concentrate here on the results in just two of these. No other differences from the Reference Case were input.

Figure 2 shows the take-up of the additional permission for housing. In zone 27 the additional permissions are fully used within 5 years; in zone 147, it takes 11 years for the additional permissions to be fully used. The difference is not that the absolute increase permitted in zone 147 is greater, but that zone 147 has generally

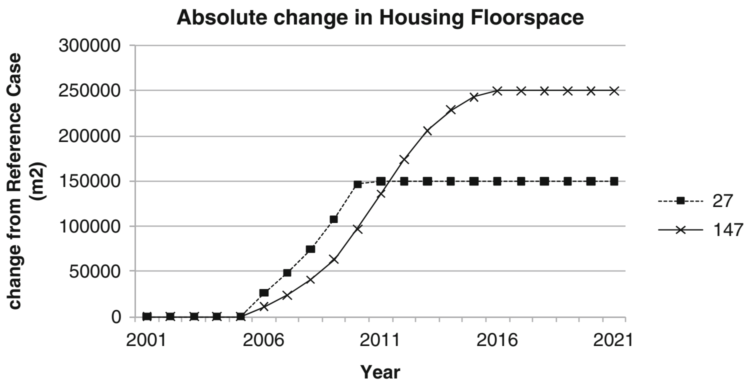


Fig. 2 Change in housing floorspace

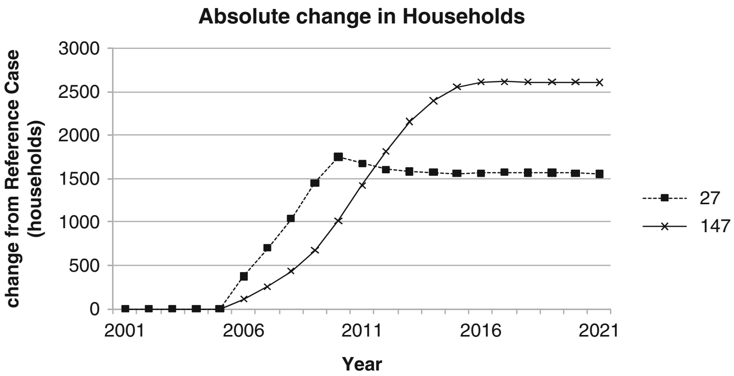


Fig. 3 Impact on households

lower rents and is a less attractive (profitable) location for housing development; it therefore attracts a lower share of developers' activity. Figure 3 shows the impacts in terms of numbers of households living in the two zones. Note that these (and all the following results) are impacts on zonal totals, not just the activity in the additional new development. In terms of households, it can be seen that the initial change in households closely follows the development of the floorspace. After 2010, however, the gain in households in zone 27 falls off somewhat; this is because there is some relocation of households from zone 27 to the continuing additional development in zone 147 (and elsewhere). Figure 4 shows the equivalent in terms of population. This shows further decline in the impact on zone 27 after the maximum impact is reached in 2010, and similarly a slightly decline in zone 147 after the

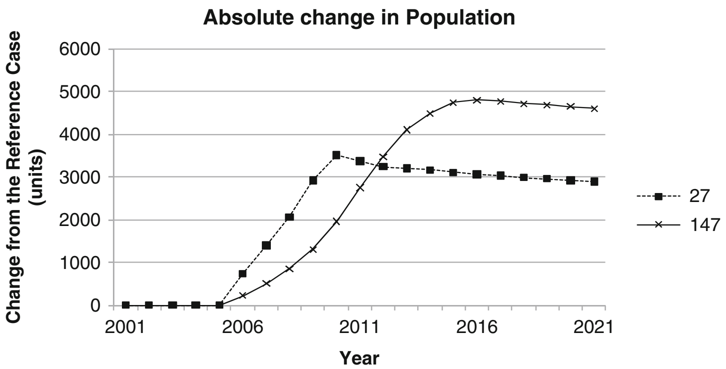


Fig. 4 Impact on population

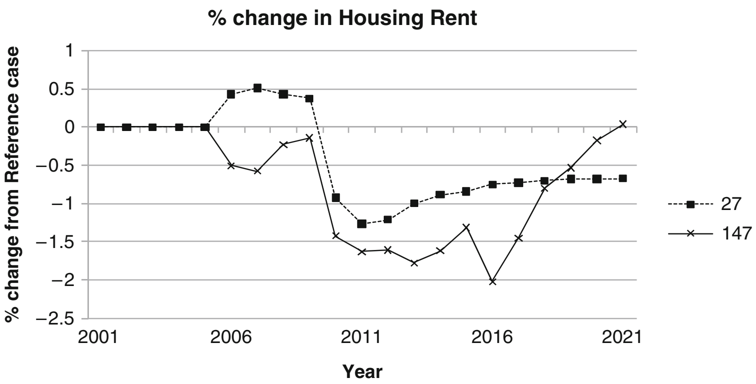


Fig. 5 Impact on housing rent

maximum positive impact in 2016. This is due to the changing mixture of households in each zone. Figure 5 shows the impacts on housing rents in these two zones.

The present model assumes a high degree of substitutability between alternative housing locations within each housing market, so an increase in supply tends to reduce rents throughout each area rather than having a highly localised impact. The curious result here is that in the first few years, the allocation of the additional permissions leads to an increase, rather than a decrease, in the housing rent in zone 27, whilst there is an immediate decrease in the rent in zone 147. This arises because zone 27 is in the Glasgow housing market area whilst zone 147 is in the Mid/South Lanarkshire area, and the effect of the additional permissions is to divert housing development from the Glasgow area to the Mid/South Lanarkshire area in the first years. This means that although the permissions for housing development in

Glasgow have increased, the supply is slightly reduced for a few years, leading to a slight and temporary increase in rents. The small scale of the rent impacts reflects the limited importance of new development compared with the very large second-hand market.

6.3 Transport Policy Impacts

We also tested the impact of a very dramatic improvement in public transport in the same corridor to the east of Glasgow. Again, some results are presented to illustrate the processes at work rather than to assess the (entirely hypothetical) transport scheme itself.

Figure 6 shows the impacts on population. It can be seen that the profiles of impact are quite different between the two zones. The gains in zone 27 reaches a peak after 5 years and then diminishes, whilst those in zone 147 grow more steadily, only slightly slowing down towards the end of the forecast. The pattern for zone 27 is quite a common one for public transport schemes, where initial impacts tend to be diluted as car ownership continues to grow and the significance of public transport is reduced. The pattern for zone 147, which relates to a much smaller initial population, is atypical because the transport scheme induces some gradual employment growth which is sufficient to produce more sustained population increases. Figures 7 and 8 show impacts on two groups within the population: children and the retired. They show that the proportional increases in numbers of children are several times greater than the proportional increases in total population. This is consistent with bus being a more important mode of transport for families with children than for other mobile households; the bus improvement attracts them disproportionately (and results in some displacement of other types of households). The decrease in the number of retired persons is an indirect consequence of this;

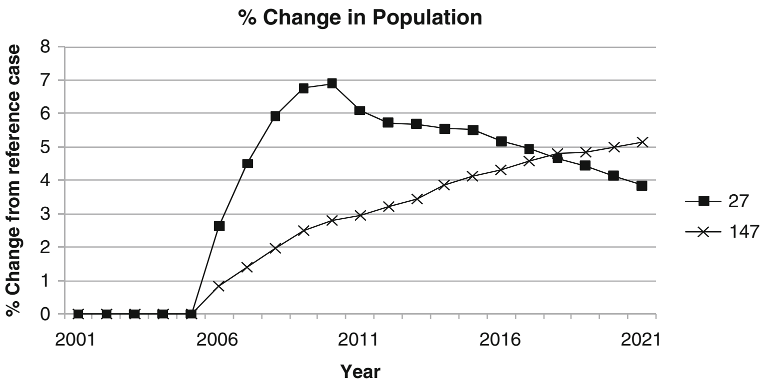


Fig. 6 Impact on population

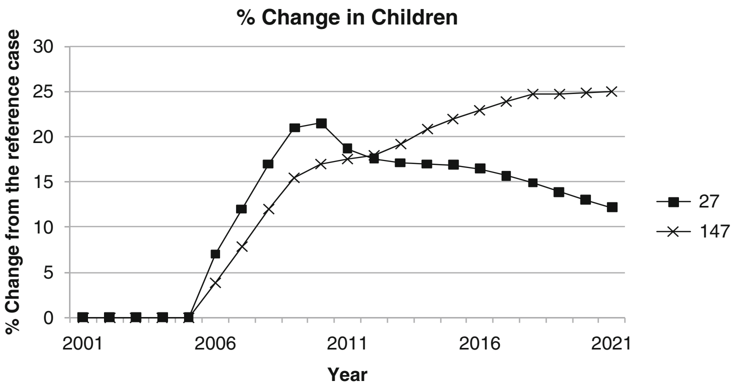


Fig. 7 Impact on children

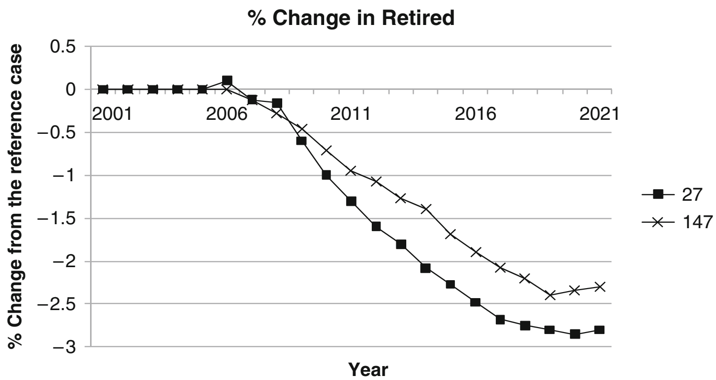


Fig. 8 Impact on the retired

families with children mostly contain adults who are some way from retirement, and hence the number of retired persons is reduced in the short- to medium-term by the decrease in the "supply" of retirees. It can be seen that decrease in the retired population is slowing or reversing towards the very end of the forecast, as this effect wears off.

7 Recent and Current Developments

The preceding sections have described the DELTA residential location model as it stood in late 2005. This section outlines a number of improvements and variations which have been pursued since then.

7.1 Treatment of Travel Costs

The reader will have noted that the money cost of travel is included in the generalised costs used in calculating accessibility measures. Changes in the cost of travel therefore affect the utility of location term through the accessibility variable, rather than by reducing the utility of consumption resulting from the expenditure of the household budget. Work is in hand to separate the money cost of travel from the rest of the accessibility measure, and to make it possible to subtract the money cost of travel from the household budget. This should have little effect on residential location itself, but will affect some of the related variables including the calculated impacts on housing rents.

7.2 Representing Different Categories of New Housing

One development now in use in several applications is to allow the user to specify that different components of the new housing supply have particular characteristics – for example, housing intended for low-income families, for young singles or couples without children, for retired persons. This is implemented by replacing the previous households and floorspace terms in the original equations (above) with a term for "expected occupiers". For existing floorspace, the expected occupiers are the same as the existing occupiers; for new floorspace, they types of expected occupiers are defined by the user to represent the different kinds of development taking place.

7.3 Modelling Short-Distance Moves

The travel-to-work areas defining the upper-level spatial units within which the residential model operates are in some cases quite large. Empirical evidence suggests that the distance-deterrent effect on local migration is strong enough to be significant within these areas, as well as being highly significant as in influence on moves between such areas, which is already taken into account in the migration model.

One recent development is to recast the residential location model in terms of explicit zone-to-zone moves. This applies to households which have an initial location, i.e. "mobile" households as defined above. The present equation for mobile households is replaced by a set of equations, each allocating "mobile" households from one "before" zone to the set of possible "after" zones, and including a distance-deterrence term varying by household type. Combining this with the "expected occupiers" function mentioned above gives the following as the main location equation for "mobile" households:

$$H(LMR)^h_{poi} = H(M)^h_{pi} \left\{ \frac{H(XA)^h_{pi}.exp(\Delta V^h_{pi})\,.\,d^h_{poi}.k^h_{pi}}{\sum_i H(XA)^h_{pi}.exp(\Delta V^h_{pi})\,.\,d^h_{poi}.k^h_{pi}} \right\},$$

where

$H(LM)^h_{poi}$	mobile households type h relocated from zone o to zone i,
$H(XA)^h_{pi}$	occupiers type h in zone i "expected" to locate in period p (calculated from floorspace and floorspace changes)
ΔV^h_{pi}	change in utility of location for households of type h locating in zone i during period p (compared with equivalent time-lagged value),
d^h_{poi}	a deterrence for households type h relocating from o to i in period p, calculated as a function of the distance from o to i,
k^h_{pi}	is a correction factor to adjust for spatial bias in the deterrence function.

"Pool" households, which by definition do not have a "before" location within the area, continue to be allocated as before.

7.4 *Other Developments*

A new version of the model is under development (as at November 2008) which will represent housing markets in considerably more detail, segmenting supply into owner occupied, privately rented and social rented sectors, and using prices rather than rents for the owner-occupied sector.

A microsimulation version of the household/individual components of the DELTA package has recently been implemented (see Feldman et al, this volume).

8 Conclusion

The DELTA package design has proved to provide a flexible context for a wide range of model developments drawing extensively on recent research results and associated thinking about urban and regional change. The residential location component is a critical element in this and one which is directly responsive to many of the other effects modelled, and indirectly linked to all the others. It is also one where it is important to take account of the dynamics of change and the ways in which significant changes in population composition – and hence in travel demand – may come about with only limited if any change in the physical stock of buildings.

We hope to continue developing the model to make better use of knowledge about residential location, residential property markets and related matters, as well as to continue using the model in our applied work.

References

Aramu A, Ash A, Dunlop J, Simmonds DC (2006) SITLUM – the Strathclyde integrated transport/land-use model. In: Proceedings of the EWGT 2006 International Joint Conferences, Politecnico di Bari

Coombe D, Skinner AJ, Simmonds DC, Davidson B (2001) Strategic environmental assessment in the Trans-Pennine corridor. Proc Inst Civil Eng Transp 147:175–190

Copley G, Skinner AJ, Simmonds DC, Laidler J (2000) Development and application of the greater manchester strategy planning model. In: Proceedings of the European Transport Conference, Transport Modelling. PTRC, London, pp 123–140

Dobson A, Simmonds DC, Capon O, Potter A, Ash A (2007) Developing a sub-regional transport strategy for regeneration and growth. Paper presented to the European Transport Conference, Noordwijkerhout, Netherlands, October 2007. Available at www.etcproceedings.org

Feldman O, Simmonds DC, Nicoll EJ (2007) The 20th Pacific Regional Science Conference, Modelling the regional economic and employment impacts of transport policies. May 2007, Vancouver

Feldman O, Mackett R, Richmond E, Simmonds DC, Zachariadis, V (this volume): A Micro-simulation Model of Household Location. In Pagliara F, Preston J, Simmonds DC (eds): Residential Location Choice: Models and Applications. Springer, Heidelberg

Hunt JD, Simmonds DC (1993) Theory and application of an iintegrated land-use and transport modelling framework. Environ Plann B 20:221–244

Minken H et al (2003) Report of the PROSPECTS project: Developing sustainable land-use and transport strategies: a methodological guidebook. Transport Economics Institute, Oslo

Nicoll EJ, Aramu A, Simmonds DC (2006) The European Transport Conference, Land-use/transport interaction modelling of the Bathgate-Airdrie railway reopening. PTRC, Strasbourg, London

Pagliara F, Preston J (2003) The impact of transport on residential location. University of Oxford Transport Studies Unit, Oxford

Simmonds DC (1994) The "Martin Centre Model" in practice: strengths and weaknesses. Environ Plann B 21:619–628

Simmonds DC (1999) The design of the DELTA land-use modelling package. Environ Plann B 26:665–684

Simmonds DC, Feldman O (2005) The ninth international Conference on computers in urban planning and urban management (CUPUM), land-use modelling with DELTA: update and experience. Available at www.cupum.org or www.davidsimmonds.com

Simmonds DC, Leitham S, White V, Williamson S (2005) Using a land-use/transport/economic model In Edinburgh, Proceedings of the European transport conference 2005, applied transport methods. PTRC, London (on CD)

Simmonds DC, Skinner A (2001) Modelling land-use/transport interaction on a regional scale: the Trans-Pennine corridor model. In: Proceedings of the European Transport Conference 2001, Applied transport methods. PTRC, London (on CD)

Simmonds DC, Skinner A (2002) The South and West Yorkshire Strategic Model. Yorkshire Humber Reg Rev 12:2

Simmonds DC, Skinner A (2004) The South and West Yorkshire strategic land-use/transport model. In: Clarke G, Stillwell J (eds) Applied GIS and spatial analysis. Wiley, Chichester

Simmonds DC, Still BG (1999) DELTA/START: adding land-use analysis to integrated transport models. In: Meersman H, van de Voorde E, Winkelmans W (eds) Transport policy. Elsevier, Amsterdam
Tinch R (1995) The valuation of environmental externalities. Report prepared for the Department of Transport
Wardman M, Bristow A, Hodgson F (1997) Valuations of noise, air quality and accessibility: evidence for households and businesses. In: Proceedings of 25th European transport forum, Seminar C: Policy, planning and sustainability. PTRC, London, pp 451–464
Whitehead TB, Simmonds DC, Preston J (2006) The effect of urban quality improvements on economic activity. J Environ Manage 80(1):1–12

The MUSSA II Land Use Auction Equilibrium Model

Francisco Martínez and Pedro Donoso

Abstract In this chapter the description of a new version of the MUSSA model is presented. The supply side of the model leading to new equilibrium problems and a solution algorithm that enhances the model performance has been significantly improved. The model is designed to forecast the expected location of agents, residents and firms, in an urban area. The model stands upon the paradigm of static market equilibrium.

1 Introduction

MUSSA is a model designed to forecast the expected location of agents, residents and firms, in the urban area, originally presented by Martínez (1996) and improved in Martínez and Donoso (2001).[1] This paper describes a new version of MUSSA model, which has significantly improved the supply side of the model leading to new equilibrium problem and a solution algorithm that enhances the model performance significantly.

The model stands upon the paradigm of static market equilibrium. The location problem assumes that real estates are allocated to the highest bidder by auctions and that market equilibrium is attained by the condition that all agents are located somewhere, therefore, supply satisfies demand. This auctioning process produces rents for each real estate in the market and simultaneously defines levels of satisfaction (benefits) to located agents at equilibrium. A discrete approach is followed for all units of demand and supply: households and firms are clustered

[1]The model is commercially available as a Windows based software since 2002 and currently distributed by Citilabs Inc. under the name of Cube Land. The software license belongs to the Chilean Government (SECTRA).

F. Martínez (✉) and P. Donoso
Department of Civil Engineering, University of Chile, Santiago, Chile
e-mail: fmartine@ing.uchile.cl

F. Pagliara et al. (eds.), *Residential Location Choice*, Advances in Spatial Science,
DOI 10.1007/978-3-642-12788-5_5,

into categories, while land is divided into zones and dwellings into types; the number of discrete units is defined by the modeller. Consumers' agents, households and firms, are assumed rational and their idiosyncratic differences are modelled by a stochastic behaviour.

The place of MUSSA in the context of other land use models can be defined from a theoretical and historical perspective. A first generation of these models was designed under the assumption that agents locate as to minimize the travel cost to other activities, which may be called the maximum access model, where the transport system has a predominant role. Several models of this class where developed following either the Alonso's (1964) bid-rent approach or the Lowry's (1964) gravity – then entropy – approach, or even a combination of these two. A second generation introduced market elements into the location problem by including rents and good prices, what we call the linear market model. Rents have been introduced in two ways, using an hedonic rent function based on average zone attraction indices, or by assuming the location options are quasi-unique so rents are the result of simulating an auction process known as the bid-auction approach. In this case, input–output tables have been used to incorporate spatially differentiated prices on goods. The third generation introduces an important amount of complexity into the model buy incorporating an explicit representation of the direct interaction between agents decisions, that is the interaction that affect behaviour in addition to the price effects. These interactions describe the fact that location options are valued, by all agents and in a significant degree, by their built environment and the location pattern, usually called zone attributes. In the economic literature (see Mas-Colell et al. 1995, p 350) this type of interactions are defined as a multilateral public externality, because it involves all agents and public or non-rivalrous goods, which we call the location externality. The significance of this phenomenon to the model formulation is that the built environment is generated by the solution of the location problem itself, then zone attributes are endogenous variables. This describes a non-linear location equilibrium problem, with a large number of endogenous variables whose solution requires more sophisticated mathematical techniques than previous generations of models. The advantage of modelling location externalities explicitly is that they describe the real inherent dynamic of the location process.

A significant difference with other land use models is that in MUSSA the interaction between consumer agents – households and firms – is explicitly described in agents' behaviour and solved to attain equilibrium. We call these interactions location externalities, which represent local attributes of a location that depends on the agents' valuation of neighbour residents and firms' valuations of agglomeration economies. These interactions are complex, so normally ignored or simplified, because they makes each agents' location choice dependant on all other agents choices, thus making the calculation of equilibrium a highly complex mathematical problem, which is solved in MUSSA by ad-hoc algorithms. It is worth noting the tremendous dynamic in the land use pattern introduced by location externalities, because each choice affects all other choices and, in theory, the whole location pattern. We shall see however, that in MUSSA this dynamic is as smooth as it is observed in reality, but it reflects a real phenomenon.

A second key difference provides a useful tool for urban planning. The model represents explicitly the whole set of physical constraints (e.g. land capacity) and planning regulations that supply must comply with. Additionally, the model allows the direct simulation of the effects of pricing incentives (taxes or subsidies) introduced by the modeller. These features provide MUSSA with tools to assess a large number of issues in urban planning, like the economic impact, i.e. social benefits/costs, of regulatory and/or pricing scenarios.

Another feature of MUSSA is that all the model parameters are calibrated by econometric methods, which provides the set of parameters required by functions that describes the behaviour of demand and supply agents. This procedure maximizes the likelihood that choices actually made by agents and observed by the modeller, are best reproduced by the set of parameters obtained conditional on the functions specified and the data used. Stated and revealed preferences data may be used. The main advantage of this methodology is that parameters can be defined as mutually consistent, considering correlation dependencies.

The improvements made in MUSSA II are significant in order to handle efficiently the non-linearity issue in the context of highly constrained space. The supply of real state units has changed, from the aggregate deterministic econometric model of the previous version to a stochastic behaviourally based logit model. Since the demand side of the original MUSSA was already based on logit models, in the new version the demand–supply equilibrium is specified as system of logit equations.

Additionally, agents' behaviour in the land use context is subject to a large number of constrains, for example households have an income, suppliers produce real state units subject to non negative profit and to comply with a large number of planning regulations. In the previous versions of MUSSA these constraints were modelled as deterministic within a constrained optimization procedure, but because of the large number of constraints this process meant an unbearable computing burden in large cities, so the solution algorithm was based on a two levels heuristic. In the new version the behaviour of agents is constrained as to comply with all constraints, such that no choice of supply or demand violates constraints. This approach is implemented in MUSSA II using the constrained logit model (Martínez et al. 2005), where a binomial logit restrains behaviour of agents to their specific feasible space.[2]

2 The Consumers' Behaviour

The goods traded in the market modelled by MUSSA II are real estate properties differentiated by type ($v \in V$) –detached, semidetached, back-to-back house, department, etc. – by the location zone ($i \in I$) and by a vector of attributes

[2]MUSSA II is specified using the RB&SM model approach of fixed points presented in Martínez and Henríquez (2007), except for the constraints on consumers' and producers' behaviour, which has been modified.

$(z = (z_{vik}, k \in K))$ –number of rooms, neighbourhood, access, land lot size, etc. The model does not restrain the dimensions $\#V$, $\#I$ or $\#K$. Consumers are households and firms classified into clusters $(h \in H)$, households are classified by socioeconomic characteristics and firms by industry. Real estate suppliers are also classified into clusters $(j \in J)$ according to their differences in production costs.

A fundamental theoretical assumption in MUSSA is that a location in the urban context is a highly scarce resource because the right to use it (by renting or buying) provides restricted access to enjoy the neighbour amenities generated by the built and natural environment. This makes each location a quasi-unique or a differentiable good, which yields a monopoly power to the landowner who obtains maximum benefit by an auction process that extract the maximum willingness-to-pay from consumers, as proposed by Alonso (1964). Consumers play in the auction game by making bids for location options, where bids represent their willingness-to-pay. Since Solow (1973) and Rosen (1974), the willingness-to-pay is a function analytically obtained as the inverse in land rents of the correspondent indirect utility function, conditional on the location choice.

We denote by V_{hvi} the indirect utility function conditional on the location option vi. Assuming that each agent "consumes" only one location and has an income y_h, this function can be expressed as $V_{hvi} = V_h(y_h - r_{vi}, p, z_{vi})$. Then, the willingness-to-pay or bid function, conditional on obtaining a given utility level U_h, is:

$$B_{hvi} \equiv I_h - f_h(p, z_{vi}, U_h), \tag{1}$$

which represents the maximum value the agent is willing-to-pay for a location described by z_{vi}, to obtain a utility level U_h given the exogenous y_h and prices p. One can understand this function in the context of choice models by thinking that the agent considers a constant utility level (taken from market conditions), and assesses her/his monetary value for each available location option in the city using this bid function. Thus, it represents the price that would make the agent indifferent on choosing any alternative location, since the utility level is assumed fixed across space. An important observation is that from (1), $e_h = B_{hvi} + f_{hvi}$ represents the expenditure function in all goods plus location cost if the consumer pays B; this is relevant for evaluating different land use patterns and it is explained in Martínez (2003). Another observation is that similar bid functions can be derived for firms directly from their differential profits obtained at different locations.

3 The Equilibrium Problem

In MUSSA II all agents maximize their individual utility (called profit for producers) subject to a set of constraints in a static context. Consumers are constrained by an exogenous income and producers by a number of regulations. The equilibrium conditions are twofold: locations are assigned to the best bidder by an absent auctioneer, and all households are located somewhere, with total demand equal to total supply of real estate options in the city.

The economic static equilibrium problem analyzed in MUSSA II is the following:

$$CP)\quad \begin{array}{l} \underset{v,i}{Max}\ (B_{hvi} - r_{vi}) \quad \forall h \\ s.t.\ \ r_{vi} \leqslant y_h - px \end{array} \qquad PP)\quad \begin{array}{l} \underset{S_{vij}}{Max}\quad S_{vij}\ (r_{vi} - c_{vij}) \quad \forall j \\ s.t.\ \ \sum\limits_{j} S_{vij} \in R_i \end{array} \quad ,$$

$$EQ1)\ l_{hvi} = \arg\max_{h \in H}(B_{hvi}) \quad \forall v, i \quad EQ2) \sum_{v \in V, i \in I} S_{vi}\ l_{hvi} = \bar{N}_h \quad \forall h.$$

The CP problem represents the consumer h's problem, which chooses a location that maximizes surplus defined by the difference between the annual willingness to pay, or bid, (B_h), minus rents (r_{vi}), subject to an available budget $y_h - px$. Here px is the expenditure in a set of goods x and p is the vector of goods' prices, assumed exogenous in the model.[3] The real estate annual rent is exogenous for the consumer but endogenous to the equilibrium yield from an auction.

The PP problem represents the supplier's j behaviour, who decides how many real estate options to offer in each submarket (v,i), denoted S_{vi}, by maximizing profit. Profits are equal to the annual rent charged minus the building plus maintenance annualized costs (c_{vij}). The supply in each zone i is subject to a set of regulations $R_i = (R_{im}, m \in M_i)$, which makes all suppliers' behaviour interdependent.

The first equilibrium condition (EQ1) represents the auction. Consumers' bids at each location (B_{hvi}) are presented to the auctioneer who assigns each allocation to the best bidder.[4] Thus, l_{hvi} is one is h is the best bidder in location (v,i), zero otherwise. This best bidder rule is sufficient to assure simultaneously that suppliers maximize profit from available supply and consumer agents maximize utility or consumers' surplus (Martínez 1992, 2000). Thus, the location that satisfies condition EQ1 implies that CP is maximized for all consumers and that real estate suppliers obtain the highest rent for their stock.

The second equilibrium condition (EQ2) relates to the whole market. Unlike markets of products where consumers decide how much –if any – they buy of each good; in this market we assume that all agents consume only one, but also not less than one location. This means that every agent has to be located somewhere at equilibrium, provided that there are sufficient location options.

---

[3]Note that, theoretically, the consumption vector x is optimal if the willingness to pay function is derived from the indirect utility function conditional on location choice.

[4]In the urban land market building properties usually have known common values, for example provided by real estate agents, so we expect that the auctioneer receives several similar bids, nevertheless, inevitably the final value is only defined by the auction. On the issue of auctions with common values see for example the review by McAffe and McMillan (1987).

The equilibrium yields the land use described by two vectors. The allocation of agents at alternative real estate options, denoted by vector $N = (N_{hvi}, h \in H, v \in V, i \in I)$, such that $\sum_{v,i} N_{hvi} = \bar{N}_h$, where $\bar{N}_h$[5] denotes the number of agents in cluster h exogenously defined; and by the vector of real estate supply $S = (S_{vi}, v \in V, i \in I)$, such that $\sum_{v,i} S_{vi} = \sum_h \bar{N}_h = \bar{N}$.

Equilibrium condition (EQ2) imply that participants in auctions outbid each other on real estate auctions, up to an equilibrium state that defines the maximum utility level attainable by each consumer in the market, represented by $U_h = U_h^*$ in (1), which yields bids at equilibrium.

The introduction of externalities represents a phenomenon that induces inherent instability in the model outcomes. This phenomenon has been widely described in social sciences (see Schelling 1978) and it is well recognized in game theory that it leads to complex non-linear mathematical formulations: small changes in initial conditions may cause dramatic differences in the location pattern and rents. Thus, it is important to note that bids functions theoretically embed location externalities, that is, the interaction between activities, by means of vector z in (1). Because the allocation of residents (neighbourhood quality) and firms (agglomeration economies) define neighbourhoods' attributes, then $z = z(N,S)$, it follows that this interaction is analytically represented by:

$$B_{hvi} \equiv I_h - f_h(p, z_{vi}(N,S), U_h) \tag{2}$$

This dependency represents a technological externality between agents, defined directly in their utility function, which operates in the urban system in addition to the pecuniary interaction through land rents.

The equilibrium presented above leads to a complex model of the city economics, extremely difficult to use for predictions. First, the facts that supply is discrete (zone system) and differentiable (location externalities), makes the location equilibrium problem mathematically untreatable for large cities in the deterministic context presented. Mathematically, it generates a complex non-linear fixed-point problem that describes a relevant and real urban dynamics introduced by the explicit representation of location externalities.

In MUSSA II, the approach to tackle this difficulty is to introduce continuity by transforming the deterministic problem into its probabilistic equivalent, which smoothes out the discontinuity associated to the agents' choice process. This approach has also the advantage of a more realistic model because idiosyncratic differences among agents within a cluster are represented by stochastic behaviour. The idiosyncratic variability on agents' behaviour takes into account the usual socio-economic and cultural differences of consumers considered in random utility theory and the variability in information and speculative behaviour in auction

[5]Overlined variables denote exogenous information required by the model.

processes; additionally, the idiosyncratic behaviour of real state suppliers describes differences in information and on strategic behaviour. The model uses the Gumbel distribution for consumers' and suppliers' behaviour, because it provides several relevant properties that help to solve the equilibrium problem.

4 The Probabilistic Bid-Auction Sub-model

In MUSSAII we conveniently assume the following bid function $B_{hvi} = I_h - f^1(U_h) - f^2(z_{vi})$, with good prices assumed as fix parameters and function f_h in (1) assumed additive. This latter assumption implies that the underpinning utility function is quasi-linear, which imposes relevant theoretical constraints into the model but with limited effect in real contexts.[6] But, it also introduces significant benefits in calculating the equilibrium, allowing the model to deal with the complex non-linearities. The assumption yields:

$$B_{hvi} = b_h^1 + b_{hvi}^2(N, S) + b^3 \tag{3}$$

where the bid function components are: b_h^1: adjusts utility U_h levels to attain equilibrium. b_{hvi}^2: describes the valuation of property attributes. Some attributes are exogenous to the location and land use distribution, like rivers, parks, hills, etc., and then they are represented by zone attractive parameters in this term. The complex attributes are those endogenous, which describe location externalities and are defined by two types of variables. First the distribution of agents' clusters in the zone, given by $N_{\bullet\bullet i}$[7] that describes attributes like neighbourhood quality by combining the characteristics of agents located in the zone (all building types) with the number of agents there located. Second, the building stock supplied in the zone (all buildings). $\overrightarrow{S}_{\bullet i}$, which describes the building environment in the zone. b^3: is a term independent of consumers and supply options, which adjusts bids to absolute levels in the whole market. This component is relevant only in the calculation of absolute value of rents and bids.

In the case of firms (non-residential activities), their bid function is derived from the profit function for each economic sector or industry. In this case, it is also assumed that the bid function is additive, like in (3).

In order to include idiosyncratic variability among consumers within a cluster, bids are assumed to be random variables: $\tilde{B}_{hvi} = B_{hvi} + \varepsilon_{hvi}$, with random terms ε_{hvi} distributed Gumbel, identical and independent (IID), justified by Ellickson (1981). From these assumptions, the (multinomial) probability that one of the $\bar{N}_h$ agents

---

[6]It merely requires that the utility function is linear in at least one good of the consumption bundle.

[7]Notation: $x_{\bullet k}$ denotes the vector of all elements of x whose second component is k.

type h is the highest bidder in (v,i), is yield – conditional on the real estate option being available – by:

$$P_{h/vi} = \frac{\bar{N}_h \exp(\mu B_{hvi})}{\sum\limits_{g \in H} \bar{N}_g \exp(\mu B_{gvi})}, \tag{4}$$

where the parameter μ is inversely proportional to the variance of the bids. Here the aggregated version of the multinomial logit probability is utilized, which includes the correction for different sizes between agents' clusters, as proposed by McFadden (1978). Thus, the expected number of agents h located at (v,i) is given by $N_{hvi} = S_{vi} P_{h/vi}$

Then, using (3) in (4), the demand model is:

$$N_{hvi} = S_{vi} \frac{\overline{N}_h \exp(\mu(b_h^1 + b_{hvi}^2(N_{\bullet\bullet i}, S_{\bullet i})))}{\sum\limits_{g \in H} \overline{N}_g \exp(\mu(b_g^1 + b_{gvi}^2(N_{\bullet\bullet i}, S_{\bullet i})))}, \tag{5}$$

where b^3 is cancelled out. In a synthetic form this is written as:

$$N_{hvi} = N(b_{\bullet}^1, N_{\bullet\bullet i}, S_{\bullet i}) \, \forall h, v, i. \tag{6}$$

This equation represents the *location fixed point*, with the probability variable both in the right and left hands of an unsolvable equation. It mathematically describes the interdependence between consumer decisions, i.e. location externalities, in which the location of an agent depends on locations of other agents (households and firms) in the same zone.

Additionally, as a direct result of the auction, the rent of a real estate (v,i) is determined by the expected value of the highest bid, which thanks to the Gumbel distribution is the known logsum or – implicit value function – given by:

$$r_{vi} = \frac{1}{\mu} \ln\left(\sum_{g \in H} \overline{N}_g \exp(\mu B_{gvi}) \right) + \frac{\gamma}{\mu} \tag{7}$$

with γ the Euler's constant. Notice that the rent depends on bids B_{hvi} and these in turn on the all other variables.

Equation (5) represents the solution of the consumer's maximizing problem (CP) and the auction condition (EQ1) simultaneously in a stochastic context. The solution of the fixed-point (5) yields the agents spatial location pattern conditional on two state variables: consumers' utilities (b^1) and supply (S). Note that the complexity of externalities remains in (5) and (5), but unlike the deterministic case discusses above, the fixed-point problem these equations is treatable, what is analyzed in detail by Martínez and Henríquez (2007).

5 Real Estate Supply Sub-model

The behaviour of real estate suppliers is twofold. First, they seek to obtain the maximum rents of their real estate stock. Second, for new development, they decide what combination of building and zone (v,i) would generate the maximum profit, subject to prevailing market regulations and rents. The first condition is already fulfilled by the auction mechanism.

There are some theoretical aspects to discuss. An important feature of the supply market is that it is highly regulated by zoning regulations, affecting both zone and building type, hence it is plausible that profit may be different by sub-markets defined by (v,i). A second issue is the heterogeneity of the suppliers, which occurs when suppliers have different profit functions. This function may be different depending on various sources of heterogeneity, for example the size of the firm that may imply different access to technology and generate different supply costs.

Another theoretical aspect is the potential for the presence of scale economies, within the same firm and zone sub-market, or scope economies (or diseconomies) across firms and zones. The most general case includes full interdependency reflected in costs functions denoted as $c_{vij} = c_j(S_{\bullet\bullet})$, where production cost depends on what is supplied everywhere by every developer. Less complex interdependencies are of course likely to occur in real markets. In any case, the supplier must determine the optimum amount to produce in each sub-market (v,i), which requires that he/she has to define an optimal vector ($S_{\cdot\cdot j}$).

In order to develop an operational supply model notice that rents are random variables, hence profits are also random. Moreover, by the property of conservation of the Gumbel distribution under maximization, rents are random variables with a Gumbel distribution that preserves the same scale parameter μ of the bids functions defined above. Thus, assuming that costs are deterministic, profits would be Gumbel distributed IID with the same scale factor as the bids. However, MUSSA II assumes that profits have a different scale parameter λ, thus allowing a more flexible model that adjusts better to the real behaviour.

In MUSSA II Hence, the expected number of residential supply units type (v,i), S_{vi}, is given by the aggregated production of each developer, which is given by the developer share of total production (S_j) times the probability that this unit type is the maximum profit option for that developer:$S_{vi} = \sum S_j P_{vi/j}$. Assuming the market share given by P_j, and the conditional probability $P^j_{vi/j}$ a multinomial logit, then:

$$S_{vi} = \bar{N} \sum_j P_j \frac{\exp(\lambda(r_{vi} - c_{vij}))}{\sum\limits_{v'i'j} \exp(\lambda(r_{v'i'} - c_{v'i'j}))}, \tag{8}$$

where λ is inversely proportional to the profit variance and $\bar{N}$ is the total number of supply units in the urban area, which is given exogenously by the total number of agents demanding locations.

Notice that, from the rent equation (7) it can be seen that b^3 is cancelled out in (8). More importantly, rents are functions of bids, then of vector N and S, and scale and scope economies make costs a function of S; therefore, the reduced form of the supply model is:

$$S_{vi} = S(b^1_\bullet, N_{\bullet\bullet\bullet}, S_{\bullet\bullet}), \tag{9}$$

which represents the *fixed-point equation* of the non-linear supply model. Equation (8) represents the solution of the producer's maximizing problem (PP) in a stochastic context and without regulations. The solution of the fixed-point (9) yields the spatial distribution of real estate supply conditional on two state variables: consumers' utilities (b^1) and consumers' allocation (N).

Notice that, despite the fact that the fixed-point problems (6) and (9) are written upon multinomial logit probabilities, they different because the state variables N and S in (8) are embedded in the rent logsum function (additionally, S is argument of the cost function). Thus, (9) has a more complex functional form than (6). Nevertheless, again the probability approach makes the fixed-point problem (8 and 9) treatable (see Martínez and Henríquez 2007).

6 Equilibrium

Here we study the supply–demand auction equilibrium (EQ2). This condition in the stochastic context of our model is expressed by:

$$\sum_{v \in V, i \in I} S_{vi} P_{h/vi} = \overline{N}_h \qquad \forall h \tag{10}$$

in which equilibrium is verified for each consumer category h and for all of them simultaneously. This condition is met if b^1 verifies that:

$$b^1_h = -\frac{1}{\mu} \ln\left(\frac{1}{\overline{N}_h} \sum_{vi} S_{vi} \exp(\mu(b^2_{hvi} - r_{vi}))\right), \tag{11}$$

which is obtained solving (11) for b^1_h. As $b^2(N,S)$ and $r_{vi}(b^1,N,S)$, this equation can be written in a reduced form as:

$$b^1_h = b_h(b^1_\bullet, P_{\bullet/\bullet\bullet}, S_{\bullet\bullet}) \tag{12}$$

and constitutes another fixed point, this time in vector b^1, whose solution verifies equilibrium conditions.

The adjustment of b^1 represents de adjustment of the utility levels that yield equilibrium: Under the additive assumption on bids, b^1 is negatively related with utility: the higher the bid for a location the lower the utility obtained (all location attributes held constant). Then the values obtained from (11) represent an index of the utilities attained by agents' clusters at equilibrium. As expected, ceteris paribus and neglecting second order effects caused by non-linearities, this index increases with $\overline{N}_h$, so utility decreases with population because supply is more demanded; more supply increases utility while higher rents have the opposite effect.

7 Modelling Constraints on Behaviour

The above models do not yet include the constraints on agents' behaviour: budget for consumers and zone regulations (R) for suppliers. The number of constrains that define a feasible domain for equilibrium variables are very large in real cities, which constitutes perhaps the highest computational burden for operational models.

MUSSA II uses a novel technique to model all constraints in the system with high computing performance. The approach is based on the Constrained Multinomial Logit Model (CML), proposed by Martínez et al. (2005). The approach consist on introducing a cut-off factor in the behaviour function, bid or profits, which has the property of making these functions tending to minus infinity as long as the attribute (or attributes) makes the constraint active. The new behaviour function is called the constrained behaviour function.

To resume how this technique works, consider for example the constrain on bids of the CP problem, which restrain bids to be positive and not greater than available income. We define the constrained bid as:

$$\tilde{B}_{hvi} = B_{hvi} + \frac{1}{\mu} \ln \varphi_{ni} + \varepsilon_{ni} \tag{13}$$

with B the unconstrained bid function. We also define φ, a cut-off factor that makes bids beyond the feasible range to have an extremely high negative value, thus reducing the probability of making that bid the best bid in the auction. Replacing the unconstrained by the constrained bid function in (4), yields the feasible probability:

$$\tilde{P}_{h/vi} = \frac{\bar{N}_h \phi_{hvi} \exp(\mu B_{hvi})}{\sum\limits_{g \in H} \bar{N}_g \phi_{gvi} \exp(\mu B_{gvi})}. \tag{14}$$

This new probability function approach zero when is evaluated at bids values out of the bid domain.

Furthermore, it is also possible to represent a set of constraints. For that the cut-off factor may be composed by a large number of factors that confine state variables to live in a multidimensional domain, with upper and lower bounds. That is $\varphi_{hvi} = \prod\limits_{k=1}^{K} \varphi_{hvik}^{L} \cdot \varphi_{nvik}^{U}$, with K the number of regulations applying to choosier h

on options v,i. This extended approach zero whenever one constraint is hit and it is useful for example to accommodate the large number of zone regulations that normally apply to each zone, representing upper and lower bounds.

In MUSSA II the cut-off factor is defined as a binomial logit function. For example, the mentioned constraint on budget represents an upper bound for rents affordable for the consumer. In our example the cut-off factor is defined by:

$$\varphi^U_{hvi} = \frac{1}{1 + \exp(\omega_k(B_{hvi} - y_h + \rho))} \begin{cases} \to 1 & \textit{if} \quad (B_{hvi} - y_h) \to -\infty \\ = \eta & \textit{if} \quad B_{hvi} = y_h \\ \to 0 & \textit{if} \quad (B_{hvi} - y_h) \to +\infty \end{cases}, \tag{15}$$

where the parameter ω defines the speed of decay of the choice probability (14) as rents approach budget; ρ and $\eta$are factors that defines the tolerance allowed by the modeller for violating the bound. Note that this tolerance is strictly positive, because binomial probabilities are positive and non-zero (except at infinite). This implies that constraints are complied only up to a minimum probability value η, which is consistent with the theory of stochastic behaviour since bounds are always subject to the agent's perception and choice on whether to comply or not.

Modelling zoning regulations is a fundamental feature of a land use model, especially in order to make it applicable as a design tool for zoning plans. The cut-off approach allows the analysis of linear and non-linear constraints, without limiting the number of regulations included. More details on how the cut-off technique is formulated and applied can be seen in Martínez et al. (2005).

Thus, using appropriate cut-off functions to the above logit functions (5) and (8) yields consumers' locations ($\tilde{N}$) that comply with the budget constraints and real estate supply ($\tilde{S}$) that comply with planning regulations. Once rents are calculated using constrained bids ($\tilde{B}$), they also internalize the effect of all constraints, both on consumers' and suppliers' behaviour. It is worth noting that, following Martínez et al. (2005), it is possible to isolate the impact of each individual regulation on rents, which represents its economic value in the urban market; thus, it allows us to make an economic assessment for each regulation.

8 The Equilibrium Equation Systems

The MUSSA II's equilibrium is represented by the simultaneous solution of the previous set of equations, which together can be written like a macro fixed-point problem such as:

$$\begin{aligned} \tilde{N}_{h/vi} &= N_{h/vi}(b^1_\bullet, \tilde{N}_{\bullet/\bullet i}, \tilde{S}_{\bullet i}) && \forall h, v, i, \\ \tilde{S}_{vi} &= S_{vi}(b^1_\bullet, \tilde{N}_{\bullet/\bullet\bullet}, \tilde{S}_{\bullet\bullet}) && \forall v, i, \\ b^1_h &= b_h(b^1_\bullet, \tilde{N}_{\bullet/\bullet\bullet}, \tilde{S}_{\bullet\bullet}) && \forall h, \end{aligned} \tag{16}$$

which is a system of dependent non-linear equations with dimension [(#H+1)(#V#I)+(#H)], with the same number of unknown variables. This system yields the solution vector $(b^1*, \tilde{N}*, \tilde{S}*)$, that is, the location of consumers, supply of real state units and relative values for rents and bids; both $\tilde{N}*$ and $\tilde{S}*$ comply with all constrains in the system. Note that each fix point is associated with either, interaction between agents (externalities the first one and economies of scale and scope the second one), or a market clearance condition (the third one).

The solution for equilibrium exists if the logit scale parameters μ and λ belong to a real range defined by Martínez and Henríquez (2007). They show that uniqueness of the solution is guarantied is scale parameters μ and λ are sufficiently low, i.e. that bids and profits do have a minimum idiosyncratic variability. Extensive numerical search indicates that such variability is usually present in real studies. When such minimum variability condition does not hold, the solution depends on the initial point because in this is the case agents' behaviour tends to be deterministic and probabilities tend to 0 or 1. In real cases studied dispersion is high enough to guarantee uniqueness.

Outputs may be used to perform rigorous economic assessments of land use scenarios. Benefits may be calculated using income compensated variations as the variation of bid values associated with scenario's changes (Martínez 2003). Thus, MUSSA II is a useful tool to obtain economic assessment of planning options such as: land use regulatory schemes, location pricing policies, transport-accessibility projects, etc. A particularly interesting application is to assess the social benefit (or cost) of each planning regulation, therefore dressing the planning process with an economic viewpoint. It can also be used by the private sector to assess expected profits from real state investments, land acquisition, location of retail and services, etc.

Some other remarks. First, it may seem that all agents, residents and non residents are allocated at each forecasting period without dependency to the previous period allocation. This interpretation is wrong, since it is possible to write the probability distributions as an incremental multinomial logit (similarly for rents), which makes evident the inter-temporal dependency in the model. Second, land use variables are endogenously updated within the equilibrium algorithm, which makes that zone attraction attributes are endogenous to the model and modified in each forecasting period. Third, residents' location behaviour consider multiple attributes, including access to most relevant activities (work, education, shopping, etc.), which are defined and calculated as transport users benefits. The trade off between attributes is defined through the bid functions by their calibrated parameters.

9 Application

The implementation of these algorithms imposes the additional challenge of obtaining a solution for a case of large dimensions. The model was applied to Santiago city, with 409 zones, 12 building types, 65 household clusters, the latter categorized by income, family size and car ownership and five firm types. This generates probability matrices with 145,000 elements that are updated in the fixed-point problem.

Additionally, the model handles 150,000 regulation constraints. The running time was 40 min, in a computer with 1.4 GHz and 1 Gb Ram memory. Therefore, despite the complexity of the interaction involved, including externalities among consumer's agents, economies of scale and scope in real state production, equilibrium conditions and regulation constraints, the model performs highly efficiently.

The model is calibrated using observed cross-section data of residential and non-residential location, rents and accessibility indices. It is also possible to use stated preference data or a mix with stated and revealed preferences. The methodology may be the standard maximum likelihood method for logit models. It is also possible to calibrate the model considering all the system equations simultaneously. The best practice will depend on the quality of the data associated to rents, consumers' location and real estate supply.

The model forecasts for any future year the urban equilibrium. The basic output includes:

- Property monthly rents, by building type and zone
- Location distribution of agents or the land use pattern, by cluster
- Buildings' distribution by zone, including houses of different sizes and types and building blocks by height levels, which defines density and average heights
- Benefits by agent clusters, which define distribution of welfare across agents
- Planning regulations and their slackness, and an index of their impact on rents

The model requires the following inputs for each forecasting year:

- Accessibility indexes by zone and preferably by cluster.
- Total city households' population by cluster and total activity for non-residential agents, also by cluster.
- The initial observation of the land use variables to initiate attributes of the built environment (by zone): average income of residents, commercial, education and services floor space.
- The initial observed distribution of real estate supply.
- The set of planning regulations and pricing (tax/subsidy) policies.
- The variables' set that describes dwelling types, e.g. lot size, floor space and building type (house or building).

The theoretical formulation of MUSSA II based on fixed point problems has also been used as component in two new tolls for planning. One tool performs optimization of regulation schemes and subsidies, which is presented in Martínez and Aguila (2004). In this case, the modeller provides the objective function and the model searches the optimal combination of regulations and subsidies. The technique was first developed to maximize social benefits in the land use market and then applied to minimize travel CO_2 emissions for Santiago city (Donoso et al. 2006).

Another tool is a dynamic model, where time is made explicit in the model as delays in the production process; the model is discussed in Martínez and Hurtubia (2006). Here auctions perform as in MUSSA II, but supply is assumed to take some time to be actually offered after the building decision is made; moreover, suppliers decide facing uncertain future. In this model excess of supply performs cycles because building is made in lump sums, which is the typical profile observed in stock economics.

Acknowledgements The authors thank MUSSA team members R Henríquez, R. Hurtubia and F Aguila, the Chilean Government (SECTRA) for allowing the publication of this paper. This paper was partially funded by FONDECYT 1060788 and Millennium Institute of Complex Engineering Systems.

References

Alonso W (1964) Location and land use. Harvard University Press, Cambridge

Donoso P, Martínez F, Zegras C (2006) The Kyoto protocol and sustainable cities: the potential use of the clean development mechanism in structuring cities for "carbon-efficient" transport. Transp Res Board 1983:158–166

Ellickson B (1981) An alternative test of the hedonic theory of housing markets. J Urban Econ 9:56–79

Lowry IS (1964) A model of metropolis. Rand Corporation, Santa Monica

Martínez FJ (1992) The Bid-choice land use model: an integrated economic framework. Environ Plann A 24:871–885

Martínez FJ (1996) MUSSA: A Land Use Model for Santiago City. Transportation Research Record 1552:126–134

Martínez FJ (2000) Towards a land use and transport interaction framework. In: Hensher D, Button K (eds) Handbooks in transport – Handbook I: transport modelling. Elsevier, The Hague, pp 145–164

Martínez FJ (2003) The effects of location externalities on social values of land and infrastructure provision. In: Button K and Hensher D (eds) Handbooks in Transport – Handbook 4 Transport and the Environment. Elsevier, Amsterdam

Martínez F, Aguila F (2004) An optimal urban planning tool: subsidies and regulations. Proceedings of the 10th World conference on transportation research, Istanbul, July 2004

Martínez F, Donoso P (2001) Modeling land use planning effects: zone regulations and subsidies. In: Hensher D (ed) Travel behaviour research, the leading edge. Pergamon, Amsterdam, pp 647–658

Martínez F, Henríquez R (2007) The RB&SM: a random bidding and supply land use equilibrium model. Transp Res B 41:631–651

Martínez F, Hurtubia F (2006) Dynamic model for the simulation of equilibrium states in the land use market. Netw Spat Econ 6:55–73

Martínez F, Aguila F, Hurtubia R (2005) The constrained multinomial logit model: a semi-compensatory choice model. Proceedings of the 8th international conference on computers in urban planning and urban management (CUPUM), London

Mas-Colell A, Whinston MD, Green JR (1995) Microeconomic theory. Oxford University Press, New York

McAffe P, McMillan J (1987) Auctions and bidding. J Econ Lit XXV:699–738

McFadden DL (1978) Modeling the choice of residential location. In: Karlqvist A et al (eds) Spatial interaction theory and planning models. North Holland, Amsterdam, pp 75–96

Rosen S (1974) Hedonic prices and implicit markets: product differentiation in pure competition. J Polit Econ 62:34–55

Schelling TC (1978) Micromotives and macrobehaviour. Norton, New York

Solow RM (1973) On equilibrium models of urban location. In: Essays in Modern Economics. Parkin M. (ed), Barnes and Nobles, New York, pp 2–16

The Impact of Transport Policy on Residential Location

Francesca Pagliara, John Preston and Jae Hong Kim

Abstract The objectives of this chapter are to assess the extent to which transport impacts on residential location decisions and hence on housing occupancy rates and house prices and to assess the extent to which transport policy decisions (such as road user charging, changes to fuel duties or the provision of light rapid transit systems) affect housing markets. This was achieved by undertaking two Stated Preference (SP) experiments in the Greater Oxford area. The aim of these experiments was to determine the key transport and location factors that householders take into account when determining their residential location. These surveys suggested that householders place high values on transport times and costs but also value low density developments, access to high quality schools, low noise levels and developments in small towns/rural areas. Stated Preference data was used to develop a hedonic pricing (HP) model which suggested much lower impacts of travel time to work, housing density and school quality on house prices than the SP choice model. Nonetheless, validation tests indicated that the HP model provided more reliable forecasts of house prices than the SP model. The HP model was used to provide preliminary forecasts of the impact of transport improvements on house prices in the Greater Oxford area.

F. Pagliara (✉)
Department of Transportation Engineering, University of Naples Federico II, Naples, Italy
e-mail: fpagliar@unina.it

J. Preston
Transportation Research Group, University of Southampton, Southampton, UK

J.H. Kim
School of Social Sciences, University of Ulsan, Ulsan, South Korea

F. Pagliara et al. (eds.), *Residential Location Choice*, Advances in Spatial Science,
DOI 10.1007/978-3-642-12788-5_6,

1 Introduction

It has been long recognized that transport factors have significant impacts on residential location. Although there is also an extensive literature on the influence of transport on residential location and on house prices, most of the previous studies have been based on revealed preference data. Thus these studies could only show the changes of transport behavior and impacts on housing markets after provision of new transport system or implementation of a new transport policy. Such studies are of significance in terms of evaluation of transport policies in the particular study region. However, the results of such studies have limitations to be directly applied to other regions due to differences of transport behavior, spatial patterns and so on in each region.

In this study, we analyze the trade-off effects between transport factors and housing price using stated preference data collected from 96 recently moved households in the Greater Oxford region in the UK. The estimation results of a bid-choice model and an hedonic model are then applied to examine the impact of different transport policies (such as a road user charge, changes of fuel duty, and a new public transport system) on housing markets. This study attempts to reveal not only the characteristics of consumer's behavior on residential location choice but also provides a framework to estimate the scope of funding transport improvements through various transport policies.

This study is novel in both its methodology, using an advanced discrete choice approach to calibrate a bid choice model of residential location, and its practical application, which permits detailed analysis of the impact of changes to the transport system on the housing markets. This study analyses three major topics as follows: firstly, individual behavior of residential location choice in terms of the trade-offs between transport factors and the others is analyzed using a discrete choice model. Secondly, the impact of transport accessibility on housing price is investigated through estimations of a bid-choice model and a hedonic price model. Finally, based on the housing price estimates, the impact of various exogenous transport policies on housing value is forecasted through a preliminary simulation method. The models are validated by comparing the forecast housing values with actual housing prices. It is then applied to examine the impact of different transport policies on housing markets in general and on property values in particular. The scope of funding transport improvement is investigated as this could provide an important long-term solution to transport funding problems.

The structure of this paper is as follows. In the rest of this section, a brief review of some recent relevant literature as well as the empirical framework are provided. In Section 2, the empirical models are specified and calibrated. In Section 3, the simulation models are applied with respect to the introduction of a road user charge, a fuel duty increase, a fuel duty decrease and the introduction of a new public transport system, Guided Transit Express. In Section 4, important findings and policy implications are summarized.

1.1 Review of Previous Study

There is an extensive literature on the influence of transport on residential location and therefore on house prices. Much of it is reviewed in Pagliara and Preston (2002a). Subsequently the Royal Institution of Chartered Surveyors and the Office of the Deputy Prime Minister (RICS Policy Unit 2002) have published the results of their study on the relationships between land use, land value and public transport. This involved a review of about 150 references. The main aim was to identify and analyze how occupier demand expressed through land values and investment yields (capital value) varied according to transport provision. In addition, ways in which better understandings of the impact of transport on property values could be used in cost benefit appraisal of transport proposals were explored. Similarly, the ways in which the impacts of transport on property values could be used in appraising land use planning and urban regeneration proposals were investigated.

Key references reviewed included Walmsley and Perrett (1992) who studied and reviewed the effects of 14 rapid transit systems in the UK, France, USA and Canada. They found that in Washington D.C. homes near stations appreciated at a faster rate than similar homes further away. Similarly, the Tyne and Wear Metro (TRL 1993) was found to have a localized effect on the housing market in a few areas, where the attractiveness of housing increased and some redevelopment took place. In general, properties near the Metro gained and maintained a slightly higher value compared with properties further away. Cervero and Landis (1995) reported that evidence from California reveals some degree of capitalization benefits, which over the long run could be expected to induce clustering around rail stations. However these impacts cannot be easily generalised. Ingram (1998) reports results of experience with new subways in Montreal, San Francisco, Toronto and Washington D.C. He found a very modest effect on metropolitan development patterns. There was also some evidence of development around stations (Toronto and Washington). Similarly, there is some evidence of CBD development impacts of high-speed rail. Banister and Berechman (2000) reviewed impacts of high-speed rail in Japan. Impacts were found at both the network and local levels. Network effects relate to the substantial increase in accessibility to key national and international markets. Another interesting study is the evaluation of the impacts that the London congestion charge had on property prices both inside and outside the zone (Zhang and Shing 2006). The congestion charge was introduced in February 2003 to reduce traffic levels in the centre of London. Postcode sector level property prices for sectors are investigated under the premise that the benefits of transport innovation can be captured by property prices. If housing markets are efficient, residential property prices should capture all the benefits and costs to commuters that a location offers. It is found that the gap between property price inside and outside the zone has actually reduced as a result of congestion charging. Also, after the implementation of congestion charge, the sensitivity of house prices with respect to distance from the boundary has fallen for sectors inside the zone relative to sectors outside the zone. An hedonic pricing model is estimated in the work of Debrezion et al. (2006) with the aim of analyzing the impact of railways on house prices in terms of distance to railway station, frequency of railway services and distance to the railway line. It has found out that dwellings very close to a station are on average about 25% more expensive than dwellings at a distance of 15 km or more. A doubling of frequency leads to an increase of house values of about 2.5% ranging from 3.5% for houses close to the station to 1.3% for houses further away. In the study by Hess and Almeida (2007) the impact of proximity to light rail transit stations on residential property values in Buffalo, New York is assessed, where light rail has been in service for 20 years, but population is declining and ridership is decreasing. Hedonic models are constructed of assessed value for residential properties within half a mile of 14 light rail stations and independent variables are included that describe property characteristics, neighborhood characteristics and locational amenities. The model suggests that, for homes located in the study area, every foot closer to a light rail station increases average property values by $2.31 (using geographical straight-line distance) and $0.99 (using network distance). Consequently, a home located within one-quarter of a mile radius of a light rail station can earn a premium of $1,300–3,000, or 2.5% of the city's median home value.

1.2 Empirical Framework

Traditional location theory examines the role of accessibility on house prices. It states that housing and accessibility to employment centers are jointly purchased in that those paying higher prices are compensated by the lower costs of commuting to the central business district (CBD) (So et al. 1996). This is the bid rent approach that has its origins with Alonso (1964).

An alternative choice approach, particularly associated with Anas (1982), examines the probability of an individual choosing a particular property as a function of the characteristics of that property, the characteristics of the individual/household and characteristics of the neighborhood in which the property is located, including accessibility. A stated preference (SP) model of this type has been calibrated and is detailed in Pagliara and Preston (2002b) and Pagliara et al. (2002a). Our empirical framework is to develop a bid choice model, which is a combination of the choice model and bid rent approaches. More specifically, we intend to follow the approach developed by Martinez (2000). That is, we intended to start by estimating the choice probability that house type v in zone i is bought by individual/household h, as given by (1):

$$P_{vi/h} = \frac{\exp \mu'(WP_{hvi} - r_{vi})}{\sum\limits_{v'i' \in \Omega} \exp \mu'(WP_{hv'i'} - r_{v'i'})} \tag{1}$$

with WP = Willingness to Pay, r = house price (rent).

The bid probability of house type v in zone i being bought by individual household h is then given by (2):

$$P_{h/vi} = \frac{\exp(\mu B_{hvi})}{\sum_{h'=H} \exp(\mu B_{h'vi})} = \exp \mu(WP_{hvi} - w_h - r_{vi} + \gamma), \tag{2}$$

where B is the bid, γ is the Euler's constant and w is the bidder's surplus, which in a perfectly competitive model will approximate to zero.

The potential terms in the WP function can be represented as:

$$WP_{hvi} = b^0 + b_h^1 + b_{vi}^2 + b_{hvi}^3. \tag{3}$$

Notice that, with the choice probability equation, it is impossible to calibrate constant WPs terms or those associated with attributes that depend only on (i.e. are constant across) households (b^0, b^1). Conversely, with the bid probability equation, constant parameters and linear terms on locations attributes (b^0, b^2) cannot be calibrated. Thus in both cases one can only calibrate truncated WP functions. In the case of bid probability equation (2), we complement the calibration by adding the rent equation (4):

$$r_{vi} = E\left[\underset{h \in H}{Max} \left(B_{hvi} + \varepsilon_{hi} \right) \right] = \frac{1}{\mu} \ln \sum_{h \in H} \exp[\mu(WP_{hvi} - w_h)] + \frac{\gamma}{\mu}. \tag{4}$$

Equation (4) allows us to calibrate the terms b^0 and b^2 in (3). This approach has been developed in Santiago, Chile, by Martinez (2000) by calibrating (2) and (4) jointly by several methods (sequential and simultaneous). However, the problem we encountered in Oxfordshire was that data were only readily available in a highly aggregate form. For example, our SP surveys only contained sufficient data for four household types (high/low income, work in city/elsewhere), two household locations (city/suburb) and two house types (detached/non detached). Similarly, house price data was only readily available at the postcode district level (e.g. OX1). As a result, an alternative approach based on hedonic pricing has been applied.

2 Model Calibration and Validation

2.1 Model Specification and Calibration

The starting point was the combination of the two utility functions of two different SP experiments (one considering access to work, the other considering access to shops) and then recalibration of the coefficients. The second step was that of converting the choice model into the bid-choice model in order to get a model, which can forecast the property chosen by households. All the attributes considered in the two SP experiments were important for the understanding of the choices made by residents in Oxfordshire (Pagliara et al. 2002b; Pagliara and Preston 2002a). However each experiment, used on its own, provided just limited information, thus a combination of the two was made. This might be thought of as a form of integrated choice experiment (Van de Vijvere et al. 1997).

The two different data sets were combined considering respectively the attributes in the second experiment held constant when calibrating the first experiment and the attributes of the first experiment held constant when estimating the second experiment. The specification of the SP choice model was as follows:

$$P_{vi/h} = \frac{\exp \mu(WP^*{}_{hvi} - r_{vi})}{\sum_j \exp\mu(WP^*{}_{hvj} - r_{vj})}, \tag{5}$$

where $P_{vi/h}$ is the probability of household h choosing property v in zone i, $WP^*{}_{hvi}$ is the truncated willingness-to-pay of household h for a property v in zone i.

The independent variables used for the empirical estimation of (5) are defined and explained in Table 1.

2.2 Estimation Results and Validation

The estimation results are reported in Table 2 for the full data set. All the attributes are significant and of the expected sign. House price, travel time and cost to work appear to be important factors influencing residential location choice. The negative value of the housing density dummy is justified by the fact that people prefer to live in areas where there is much open land. The negative value of the location dummy CITY means that the preference is for living away from the city i.e. in country towns and rural areas – we refer to these locations below as SUBURBAN areas. Another important factor is travel cost to shops, which is negative and significant, i.e. people prefer to live close to shopping centers. The positive and highly

Table 1 Variable definition

Variable	Definition
HPrice	The current market value of the house (in pounds)
TTWork	The total time (in minutes) spent making a single trip from the house to the workplace
TCWork	The total cost (in pence) spent making a single trip from the house to the workplace
DENS	A dummy equal to 1 if the house is in an area with no open land, 0 otherwise
CITY	The location within the boundary of Oxford City (OX1, OX2, OX3, OX4)
TCShop	The total cost (in pence) spent making a single trip from the house to a large supermarket
QSCH	A dummy, equal to 1 if the house is in an area with good schools, 0 otherwise
NOISE	A dummy equal to 1 if the house is in a noisy area, 0 otherwise
DETACH	A dummy equal to 1 if the house is detached, 0 otherwise

Table 2 Stated preference choice model estimation results

Variable	Coefficient (t-value)
HPrice	−0.328-E05 (−2.620)
TTWork	−0.449-E01 (−9.033)
TCWork	−0.660-E02 (−5.125)
DENS	−0.498 (−7.593)
CITY	−0.291 (−4.027)
TCShop	−0.321-E02 (−2.790)
QSCH	0.727 (12.849)
NOISE	−0.877 (−14.015)
DETACH	0.323 (3.653)
No. of observations	3,072
L(*)	−2,928
L(0)	−3,374
ρ^2	0.132

Table 3 Probabilities computation for persons working in city

Income	House Type	Residence	Estimated probability	Actual probability
High	Detached	City	0.33894	7.89E-02
		Suburban area	0.24377	0.28947
	Non-detached	City	0.24654	0.39474
		Suburban area	0.17074	0.23684
Low	Detached	City	–	–
		Suburban area	0.13166	6.67E-02
	Non-detached	City	0.46091	0.48889
		Suburban area	0.40744	0.44444

significant value of the quality of school dummy means that people prefer to live in areas with good schools. The negative and significant value of the noise dummy means that the choice of residence is strongly influenced by the noise level of a given area. The preference is to live in quiet areas. The positive and significant value of the detached dummy means that people prefer, all other things being equal, to live in detached houses.

Households are grouped according to household income and workplace location. Household income categories (two levels) are low (in our sample less than/equal £42.50 K per year) and high (greater than £42.50 K per year). Workplace locations are CITY, i.e. within the boundary of Oxford city (OX1, OX2, OX3, OX4) and SUBURBAN area (the remaining part), i.e. outside Oxford city (two levels). Therefore, four categories of households have been identified. Residential zones are again CITY and SUBURBAN area (two levels) and whether households live in a detached house or not (two levels). Again four categories have been identified. Therefore 16 (= 4 × 4) different segments have been identified and the estimated and actual probabilities are reported in Tables 3 and 4. The estimated probabilities were obtained by operationalising (5) with the coefficient values in Table 2. The actual probabilities came directly from our surveys.

Table 4 Probabilities computation for persons working in suburban area

Income	House type	Residence	Estimated probability	Actual probability
High	Detached	City	–	–
		Suburban area	0.28989	0.42857
	Non-detached	City	0.15119	0.28571
		Suburban area	0.55892	0.28571
Low	Detached	City	–	–
		Suburban area	0.71049	0.16667
	Non-detached	City	4.51E-02	0.16667
		Suburban area	0.19879	0.66667

For people working in the CITY (Table 3) and choosing a detached house, the actual probability of residing in a SUBURBAN area is higher for those in the high income group compared to those in the low income group. The former, thanks to their budget, can choose to live in a detached house with higher prices. Conversely, the actual probability of choosing a non-detached house and residing in a SUBURBAN area is higher for those in the low income group. For people working in the SUBURBAN area (Table 4), the actual probability of residing in the SUBURBAN area and choosing a detached house is higher for those in the high income group (although our model's forecasts do not replicate this). High income group can move to the CITY. The probability of living in the CITY and choosing a non-detached house is also higher for the high income group, reflecting their ability to afford more expensive properties.

2.3 Comparison Between Actual and Estimated Probabilities

The estimated probabilities in Tables 3 and 4 fail to take into account supply side constraints. For example, the stock of detached houses is limited, particularly in the CITY. Therefore, in order to derive a bid-rent model from our SP choice model the following procedure was developed, to reflect the aggregate nature of our data. The willingness-to-pay is related to utility by the following relation: $U_{hvi} = \mu (WP_{hvi} - r_{vi})$ and given that the log sum is the appropriate measure of expected maximum utility the following expression can be computed (for a model calibrated at the individual level):

$$b_i = \frac{1}{\theta\mu} \ln \sum_h H_h S_{hi} \exp(\theta\mu WP_{hi}), \tag{6}$$

where:

I	is the zone index,
H	is the household type (high/low income, work in city/suburban area),
H_h	is the number of households of type h,
S_{hi}	is the density of households of type h in zone i,
$\theta\mu$	are parameters to be estimated.

Given that the parameter μ has been estimated in the SP choice model the re-scaling parameter θ may be estimated from aggregate data, based on the Berkson-Theil method as follows:

$$\ln\left(\frac{P_{h/vi}}{1-P_{h/vi}}\right) = \theta\left(U_{h/vi} - U_{h/vj}\right), \tag{7}$$

where:

$P_{h/vi}$	is the probability house and location type vi is chosen by individual h,
$U_{h/vi}$	is the utility of the chosen alternative for individual h and house and type location type vi,
θ	is the parameter to be estimated.

From (7), the θ coefficient has been obtained as 0.464 with a t-statistic of 4.244. Hayashi and Doi (1989) found values for a bid MNL (Multinomial Logit) scale parameter of about one third of the choice MNL scale parameter. This result is comparable to what we have obtained. From Table 2 we can estimate the willingness to pay as equal to:

$$\begin{aligned} WP_{hvi} = (&-0.449\text{-}E01\ TTWork - 0.660\text{-}E02\ TCWork - 0.498\ DENS \\ &- 0.291\ CITY - 0.321\ TCShop + 0.727\ QSCH - 0.877\ NOISE \\ &+ 0.323\ DETACH)/0.328 - E05 \end{aligned}$$

where 0.328-E05 is the parameter value of the house price attribute. Operational models recognise that the estimation of utility or willingness-to-pay functions is subject to inaccuracy in terms of fully describing actual behavior, and best defined as a stochastic variable. Let us assume that bids are given by $B_{hvi} = B_{hvi} + \varepsilon_{hvi}$ where $B_{hvi} = WP_{hvi} - w_h$ is the deterministic component, ε_{hvi} is a random error term and w_{h} is the speculative term that has to be equal for all location options to make sure that the consumer is indifferent to any option where he or she is the best bidder. A family of location models can be proposed by assuming different distributions of the random term. One of the most applied is the Gumbel distribution (Martinez 2000). Assuming the stochastic terms as independent and identically distributed Gumbel (IIG) with a scale parameter θ, we obtained the following expression for the expected maximum bid, which directly represents the expected rent at location (vi):

$$r_{vi} = E[MAX(B_{hvi} + \varepsilon_{hi}] = \frac{1}{\theta\mu} \ln \sum_{h \in H} \exp[\theta\mu(WP_{hvi} - w_h)] + \frac{\gamma}{\theta\mu}, \tag{8}$$

where γ is the Euler's constant (approximately 0.577) and we assume $w_h = 0$.

By applying appropriate weighting parameters (H_h and S_{hi}), house price forecasts for our aggregate categories are shown in Table 5. However, the forecasts systematically under predicted actual prices, with this being particularly evident for suburban non-detached properties. Moreover, the bid-choice prices that result are very sensitive to the weighting parameters used. However, it is important to note that our SP data was collected in April 2002 whilst our actual data was provided

Table 5 Comparison between bid-choice prices and actual prices (unit: £)

Residential area	House type	Bid-choice prices	Actual prices	% Change
City	Detach	362,790	385,000	−5.8
City	Non-detach	213,152	255,532	−16.6
Suburban	Detach	296,722	293,750	−1.0
Suburban	Non-detach	66,227	174,065	−62.0

Table 6 Estimation results of Hedonic Price Model

Attribute	β	t-statistic
Constant	83.072	5.296
NBEDROOMS	29.457	11.064
TTWork	−1.264	−8.482
TCWork	−3.631	−1.975
DENS	−16.345	−4.047
TCShop	−3.749	−0.921
QSCH	12.567	3.882
NOISE	20.531	4.634
DETACH	105.979	14.137
CITY	29.302	13.023
HIGH_INCOME	54.952	13.023
WORKCITY	14.462	2.295
R^2 : 0.276		

from the Land Registry for the period July to September 2002 (from www.proviser.com). There will have been some increases in house prices between these two dates. Between mid 2000 and mid 2001 house prices in Oxford rose by 19%, whilst between mid 2001 and mid 2002 house prices rose by 37%. This suggests that land prices in our study area could have risen by as much as 12% in the period between our surveys (in April, 2002) and the mid-term date of the Land Registry data used (August, 2002). In order to reconcile results, a hedonic pricing (HP) regression has been undertaken (Rosen 1974). This assumes that house prices are some implicit (hedonic) function of a bundle of attributes. The data set used is that collected in the stated preference exercise (3,072 observations) but includes data on the existing property as well as the two hypothetical properties presented in each scenario and the option of staying with the existing property is modelled in addition to choosing the hypothetical properties. An element of revealed preference data is therefore included in the HP model. The advantages of HP methods include their links with market data and their widespread application. Disadvantages include assumptions of identical incomes and preferences for all consumers, no transaction costs and fixed supply (Tinch 2002).

A linear model was specified, with the explained variable being house price (in thousands of pounds) and the explanatory variables being the other attributes in the SP experiment but in addition the number of bedrooms was added as an attribute. Table 6 reports the estimation results of the hedonic model. Note that TCWork and TCShop are expressed in pounds rather than pence. Most parameter values are significant at the 5% level (but note the repeat observations problem will be at play here), although the TCWork parameter value is only just significant and

Table 7 Values (in £) in terms of house price of a unit change in attributes

Attribute	SP		HP	
	Attribute values	t-statistics	Attribute values	t-statistics
TTWork	−13,680	(2.67)	−11,264	(8.48)
TCWork	−2,011	(2.37)	−36.31	(1.98)
DENS	−151,868	(2.42)	−16,345	(4.05)
TCShop	−978	(2.11)	−37.49	(0.92)
QSCH	221,650	(2.47)	12,567	(3.88)
NOISE	−267,406	(2.50)	20,531	(4.63)
DETACH	98,538	(2.28)	105,979	(14.14)
CITY	−88,807	(2.31)	29,302	(5.64)

the TCShop parameter value is insignificant. The goodness of fit is only modest, suggesting this model may be affected by multicollinearity. This is a common problem with HP models and might also explain the implausible sign of the NOISE parameter value, although this could also reflect model mis-specification. For example, our model only includes local accessibility measures but locations close to motorway junctions, although having high noise levels, may also have good long distance accessibility.

Table 7 reports some marked differences in attribute values obtained from the SP and HP approaches. The SP valuations of TTWork, QSCH, and DETACH are more than 10 times greater than the values from the HP model, whilst the SP valuation of DENS is over 9 times greater than that of the HP model. Moreover, the valuations of TCWork and TCShop are even more substantially higher in the SP model than the HP model. This leads to a high value of time estimated in the HP model of almost 35 pence per minute compared to the much more plausible 8 pence per minute in the SP experiment.

The Department for Transport's Transport Economic Note (http://www.roads.dft.gov.uk/roadnetwork/heta/highway/04.htm) suggests a market price value of working time for car drivers of 35.1 pence per minute (23.3 pence per minute for all modes) and 7.5 pence per minute for non-working time (all at 1998 prices). This suggests that the HP model's values of time are broadly consistent with values of working time and the SP model's values are broadly consistent with values of non-working time. Around 13% of distance traveled and around 5% of all trips by car are in the course of work (Source: http://www.webtag.org.uk/webdocuments/3_Expert/5_Economy_Objective/3.5.6.htm#012 <Accessed 27 February 2007), and hence we would expect the "true" average values to be closer to the SP than the HP results, although the HP results may reflect marginal values.

It is also noteworthy that the CITY dummy variable has a negative impact on house prices in the SP model but a positive influence on house prices in the HP model. This may suggest that those living in the CITY are less likely to move than those living in rural areas or country towns. Similarly, the NOISE parameter has a strong negative effect on house prices in the SP model but a modest positive effect in the HP model.

The comparisons in Table 7 suggest that the values from the SP model may have serious upward biases. These are corrected in the HP model by better representing

Table 8 Comparison between hedonic prices and actual prices (unit: £)

Residential area	House type	Hedonic prices	Actual prices	% Change
City	Detach	357,662	385,000	−7.10
City	Non-detach	216,976	255,532	−15.08
Suburban	Detach	293,652	293,750	−0.03
Suburban	Non-detach	187,673	174,065	7.81

housing characteristics by including a parameter for the number of bedrooms and thus reducing specification error. It is interesting to note that the one valuation that was similar in the SP and HP models was the DETACH dummy variable. Furthermore, the inclusion of the choice of the current property may have led to a better representation of budget and other constraints and thus reduced non-commitment bias.

Table 8 presents the resultant forecast prices from the HP model. Compared to Table 5, there is no longer a systematic under valuation of prices. The main advantage of the HP approach appears to be the more accurate forecasts of suburban non-detached housing.

2.4 Application of the Hedonic Price Model to the Study Area

The final stage of the study has involved the application of the model to assess the impact of transport policy changes on residential location preferences and the resultant prices. The area under study is the corridor Kidlington–Oxford–Abingdon.

The first step has been that of applying the HP model to the relevant post code districts, computing the resultant house prices and comparing them with the house prices in the Land Registry database (www.proviser.com). The aggregate nature of this analysis should be reiterated. These computations have given the following results shown by Tables 9 and 10. The sensitivity tests provided by Table 10 involved adjusting the HP model so that the values of time are in line with those obtained in the SP model. This was done by increasing (in absolute terms) the parameter values of travel costs to work and to shop to –18.5 from –3.631 and –3.749 respectively.

The unweighted HP model slightly underestimates house prices on average, as might be expected given the recent rapid increases (estimated as being as much as 12% between April and August 2002). The last column shows the percentage change of house prices since 2002.

The re-weighted model underestimates house prices to a greater extent as might be expected as the impact of transport cost has been increased somewhat artificially. Both models are reasonably accurate in aggregate, but both versions of the model have a problem in capturing the premium attached to OX2 (North Oxford) addresses by people working at the University of Oxford.

Table 9 (Unweighted) Comparison between hedonic prices and actual prices

Area	House type	Hedonic prices (£)	Land registry prices (£) (2002)	Δ% (2002)	Land registry prices (£) (2008)	Δ% Land registry prices (2002–2008)
OX1	Detach	307,670	297,125	3.55	529,272	43,861
	Non-detach	178,524	163,571	9.14	266,428	38,605
OX2	Detach	355,240	506,549	−29.87	543,881	68,640
	Non-detach	226,093	292,953	−22.82	327,603	10,576
OX4	Detach	284,196	307,500	−7.58	426,059	27,826
	Non-detach	153,277	148,686	2.96	197,594	24,751
OX5	Detach	280,293	300,083	−6.59	426,906	29,707
	Non-detach	157,653	145,666	8.23	169,499	14,060
OX13	Detach	238,246	279,286	14.69	525,800	46,883
	Non-detach	172,211	164,055	4.97	155,000	−5,841
OX14	Detach	239,772	240,740	−0.40	356,135	32,402
	Non-detach	173,737	155,369	11.82	171,189	92,412
Average		230,576	250,147	−3.44	341,280	23,245

Table 10 (Weighted) Comparison between hedonic prices and actual prices

Area	House type	Hedonic prices (£)	Land registry prices (£) (2002)	Δ% (2002)
OX1	Detach	289,990	297,125	−2.40
	Non-detach	160,844	163,571	−1.67
OX2	Detach	340,789	506,549	−32.72
	Non-detach	215,643	292,953	−26.39
OX4	Detach	275,014	307,500	−10.56
	Non-detach	146,604	148,686	−1.40
OX5	Detach	274,280	300,083	−8.60
	Non-detach	147,556	145,666	1.30
OX13	Detach	225,907	279,286	−19.11
	Non-detach	169,330	164,055	3.22
OX14	Detach	215,974	240,740	−10.29
	Non-detach	169,397	155,369	9.03
Average		219,277	250,132	−8.30

3 Model Application and Policy Simulation

The model has been applied by examining the impact of some transport policy scenarios on the housing market and in particular on house prices. Further details are given in Pagliara et al. (2002b). Actual travel times and costs to/from each of the postcode districts have been computed for both car and bus, whilst journey to work movements have been obtained from the 1991 Census.

In order to evaluate the effect of transport policy changes on residential location preferences the incremental logit model (Kumar 1980; Preston 1991) has been applied to calculate the mode shares of travel in the before and after situation. For the simple two mode situation, the formulation is as follows:

$$P'_C = \frac{P_C \exp(U'_C - U_C)}{P_C \exp(U'_C - U_C) + P_{PT} \exp(U'_{PT} - U_{PT})}, \tag{9}$$

where:

$P'_C(P_C)$	is the proportion of people choosing car in the after (before) situation,
$U'_C(U_C)$	is the utility measure of car in the after (before) situation,
P_{PT}	is the proportion of people choosing public transport in the before situation,
$U'_{PT}(U_{PT})$	is the utility measure of public transport in the after (before) situation.

Four different scenarios have been examined for illustrative purposes. These are (1) Road User Charge, (2) Fuel Duty Increase, (3) Fuel Duty Decrease and (4) Introduction of the GTE system.

In order to calculate the proportions in the after situation with the provision of the new mode of transport (Scenario 4), the extended logit model formulation has been applied. This has the following formulation:

$$P'_{PT} = \frac{P_{PT}\left[exp(U'_{NT} - U_{XT}) + exp(U'_{NT} - U_{XT})\right]^{\phi}}{P_C\, exp(U'_C - U_C) + P_{PT}\left[exp(U'_{NT} - U_{XT}) + exp(U'_{NT} - U_{XT})\right]^{\phi}}, \tag{10}$$

where

$_{XT}$ = Existing Public Transport mode (Bus),

$_{NT}$ = New Public Transport Mode (Guided Transit Express),

ϕ = Expected Maximum Utility Parameter Value.

The utility of travel in the before situation is therefore:

$$U^* = \ln(\exp U_C + \phi \exp U_{XT}). \tag{11}$$

This can been divided through by the cost parameter in order to derive a generalised cost measure. Similarly in the after situation, the utility of travel can be determined as:

$$U'^* = \ln[\exp U'_C + \phi(\ln(\exp U'_{XT} + \exp U'_{NT}))]. \tag{12}$$

A problem with such composite utility terms is that they cannot easily be split into time and cost changes. Naïve averaging of times and costs will give implausible results. For example suppose a new public transport mode such as Guided Transit Express is introduced and which on average has the same speeds as other modes (but will be faster for some passengers) but fares are slightly higher than the bus. Averaging would indicate that overall the generalized costs of travel have gone up and overall travel demand has gone down. However, this is not plausible. Barring congestion effects, a new transport mode cannot reduce demand, at worst it can have no effect. The solution to this was to use the proportionate change in generalized costs (and its component time and cost elements) derived from (11) and (12) to adjust both travel time and travel cost. Ideally, our HP model would include the composite utility terms as explanatory variables but this was not possible with the data sets used in this project.

Equations (9) and (10) thus determine the impact on modal shares of transport policy changes. The impacts on travel times and costs are then computed from (11) and (12) and these are then fed into the HP model (Table 6) to determine the impact on house prices.

3.1 Road User Charge

In the first scenario, a cordon toll of £1 per day is introduced for those traveling into or through the city centre (specified as OX1 and OX2). This is similar to the level of road user charging being considered at the time in cities such as Bristol. It is supposed that in the city centre there is an increase in speeds equivalent to a decrease of 5 min of travel time. Tables 11 and 12 report the resultant prices in the two situations, reflecting a value of time of 35 pence per minute and 7 pence per minute respectively.

The charge of £1 to enter the central area (OX1 and OX2) from OX4, OX5, OX13, and OX14 has caused a decrease in house prices in the outer areas of between 2% and 18%, with this decrease in house price being more marked in the (re)weighted model as expected given the increase in the cost parameter value. The slight increase in house prices in OX1 and OX2 of between 1% and 6% is due to the forecast reductions of travel time in those areas. Overall, the unweighted HP model suggests an average housing price decrease of 1.7%. However, this is composed of an average 1.8% increase in house prices in the charged area and a 3.4% decrease in the non charged area. A similar pattern has recently been forecast, using a different methodology, in work in Greater Manchester undertaken as part of a parallel New Horizons project (David Simmonds Consultancy 2003).

Table 11 (Unweighted) Prices in Scenario 1 (£)

Area	House type	Before	After	% Change
OX1	Detach	307,670	311,591	1.26
	Non-detach	178,524	181,445	1.61
OX2	Detach	355,240	361,250	1.66
	Non-detach	226,093	232,218	2.64
OX4	Detach	284,196	277,080	−2.57
	Non-detach	153,277	146,110	−4.91
OX5	Detach	280,293	273,953	−2.31
	Non-detach	157,653	151,009	−4.40
OX13	Detach	238,246	232,122	−2.64
	Non-detach	172,211	165,311	−4.17
OX14	Detach	239,772	233,001	−2.91
	Non-detach	173,737	167,882	−3.49
Average		230,576	227,748	−1.69

Table 12 (Weighted) Prices in Scenario 1 (£)

Area	House type	Before	After	% Change
OX1	Detach	289,990	297,123	2.40
	Non-detach	160,844	166,444	3.36
OX2	Detach	340,789	348,122	2.11
	Non-detach	215,643	229,163	5.90
OX4	Detach	275,014	232,175	−18.45
	Non-detach	146,604	125,001	−17.28
OX5	Detach	274,280	241,745	−13.46
	Non-detach	147,556	127,122	−16.07
OX13	Detach	225,907	198,122	−14.02
	Non-detach	169,330	146,145	−15.86
OX14	Detach	215,974	188,412	−14.63
	Non-detach	169,397	207,146	18.22
Average		219,277	208,893	−6.48

It should be noted that the HP model does not take into account relative accessibility (in contrast to the bid rent and bid choice approaches) and hence might understate the differential between charged and non-charged areas. On the other hand, we have assumed that the revenue raised from road user charging is not hypothecated into, for example, improvements to local public transport or bettering the local urban environment. Such hypothecation could reduce the impact on property prices in the outer area (David Simmonds Consultancy 2003).

3.2 Fuel Duty Increase

Scenario 2 considers an increase of 10% in fuel duty assuming that the petrol cost is equal to 70 pence/l and that the fuel duty is equal to 60 pence. Tables 13 and 14 show the prices after this increase. As expected, such an increase has brought a reduction of house prices everywhere, in this case of between 2% and 13%. Again this is more marked in the reweighed model. Overall, the unweighted model suggests an average house price reduction of 3.4%.

3.3 Fuel Duty Decrease

Scenario 3 considers a decrease of 10% in fuel duty, again assuming that the petrol cost is equal to 70 pence/l and that the fuel duty is equal to 60 pence. Tables 15 and 16 show the resultant house prices after this decrease. In this scenario, house prices increase everywhere as expected by between 2% and 9%. Overall, the unweighted model suggests an average house price increase of 3.2%. However, in comparison with a fuel duty increase, some interesting asymmetries are revealed. This reflects the non-linear nature of the incremental logit models used to forecast the impact of fuel duty on mode choice and hence on overall travel times and costs.

Table 13 (Unweighted) Prices in Scenario 2 (£)

Area	House type	Before	After	% Change
OX1	Detach	307,670	300,712	−2.31
	Non-detach	178,524	173,204	−3.07
OX2	Detach	355,240	346,041	−2.66
	Non-detach	226,093	216,895	−4.24
OX4	Detach	284,196	276,686	−2.71
	Non-detach	153,277	146,488	−4.63
OX5	Detach	280,293	272,544	−2.84
	Non-detach	157,653	152,268	−3.54
OX13	Detach	238,246	230,620	−3.31
	Non-detach	172,211	164,585	−4.63
OX14	Detach	239,772	230,828	−3.87
	Non-detach	173,737	169,793	−2.32
Average		230,576	223,389	−3.35

Table 14 (Weighted) Prices in Scenario 2 (£)

Area	House type	Before	After	% Change
OX1	Detach	289,990	272,063	−6.59
	Non-detach	160,844	153,144	−5.03
OX2	Detach	340,789	330,122	−3.23
	Non-detach	215,643	201,075	−7.25
OX4	Detach	275,014	248,126	−10.84
	Non-detach	146,604	132,122	−10.96
OX5	Detach	274,280	259,122	−5.85
	Non-detach	147,556	137,896	−7.01
OX13	Detach	225,907	199,889	−13.02
	Non-detach	169,330	158,015	−7.16
OX14	Detach	215,974	203,166	−6.30
	Non-detach	169,397	158,126	−7.13
Average		219,277	204,406	−7.53

Table 15 (Unweighted) Prices in Scenario 3 (£)

Area	House type	Before	After	% Change
OX1	Detach	307,670	314,467	2.16
	Non-detach	178,524	185,110	3.56
OX2	Detach	355,240	364,179	2.45
	Non-detach	226,093	235,033	3.80
OX4	Detach	284,196	296,722	4.22
	Non-detach	153,277	160,445	4.47
OX5	Detach	280,293	286,299	2.10
	Non-detach	157,653	161,978	2.67
OX13	Detach	238,246	249,191	4.39
	Non-detach	172,211	176,156	2.24
OX14	Detach	239,772	249,376	3.85
	Non-detach	173,737	178,341	2.58
Average		230,576	238,108	3.21

Table 16 (Weighted) Prices in Scenario 3 (£)

Area	House type	Before	After	% Change
OX1	Detach	289,990	310,122	6.49
	Non-detach	160,844	175,369	8.28
OX2	Detach	340,789	358,889	5.04
	Non-detach	215,643	236,728	8.91
OX4	Detach	275,014	299,125	8.06
	Non-detach	146,604	159,223	7.93
OX5	Detach	274,280	289,693	5.32
	Non-detach	147,556	156,478	5.70
OX13	Detach	225,907	240,102	5.91
	Non-detach	169,330	178,690	5.24
OX14	Detach	215,974	233,489	7.50
	Non-detach	169,397	179,236	5.49
Average		219,277	232,179	6.66

3.4 Introduction of the GTE System

Another scenario considered is the provision of the Guided Transit Express (GTE) system. This system will serve the Kidlington–Oxford–Abingdon corridor studied for the other scenarios. The aim of our model is to estimate the impact of the system on house prices. Oxford GTE is a proposal for a guided bus way to allow an express bus network to connect the city centre to key Park and Ride sites and surrounding towns. The GTE route involves sections of off-highway (on a guide way) and on-highway alignment. The concept of a guided bus scheme to link Park and Ride sites to the north and south of Oxford with the City centre was first proposed in 1994 by the Oxford Bus Company. At that time, the project was being considered as a public– private partnership involving the County and City Councils and key private sector organizations including Oxford's bus companies, national construction companies and local business and education establishments.

The intention of the GTE is to provide fast, reliable and congestion free routes for public transport into the city centre of Oxford from country towns and villages. The scheme will enable public transport operators to provide a more effective network of routes and interchange points to serve Oxfordshire County and Oxford City (CJ Associates 2001). It will help to ease congestion by removing cars from the roads and routing some existing bus services from the roads into the guide way. This will also help to improve the quality of the city centre in terms of both air quality and pedestrian amenity. In addition, by providing additional public transport services the stress on existing networks will be reduced.

The provision of the GTE will change travel times and costs. It is supposed that this system will have fares around 10% higher than competing bus services and will reduce travel times between the areas directly served by the route. Tables 15 and 16 show the forecast house prices after the introduction of the new system.

With the provision of the GTE prices go up everywhere. The increase in house prices varies between 1% and 7%. This is not surprising since the new system brings reductions in travel times and makes all areas more accessible. The more accessible

Table 17 (Unweighted) Prices in Scenario 4 (£)

Area	House type	Before	After	% Change
OX1	Detach	307,670	320,348	3.96
	Non-detach	178,524	191,200	6.63
OX2	Detach	355,240	363,100	2.16
	Non-detach	226,093	233,299	3.09
OX4	Detach	284,196	290,602	2.20
	Non-detach	153,277	156,430	2.02
OX5	Detach	280,293	286,739	2.25
	Non-detach	157,653	164,510	4.17
OX13	Detach	238,246	246,113	3.20
	Non-detach	172,211	180,078	4.37
OX14	Detach	239,772	245,750	2.43
	Non-detach	173,737	179,660	3.30
Average		230,576	238,152	3.32

Table 18 (Weighted) Prices in Scenario 4 (£)

Area	House type	Before	After	% Change
OX1	Detach	337,035	345,831	2.61
	Non-detach	207,889	215,683	3.75
OX2	Detach	393,457	399,396	1.51
	Non-detach	264,310	269,249	1.87
OX4	Detach	313,967	319,575	1.79
	Non-detach	184,728	188,379	1.98
OX5	Detach	276,222	282,439	2.25
	Non-detach	192,497	200,238	4.02
OX13	Detach	249,749	257,711	3.19
	Non-detach	203,172	211,577	4.14
OX14	Detach	250,507	256,131	2.24
	Non-detach	203,930	210,240	3.09
Average		256,455	263,037	2.70

is an area the higher are the price increases, with the biggest increases in central Oxford (OX1), Kidlington (OX5) and Abingdon (OX13 and OX14). The unweighted model suggests an average house price increase of 3.2%. This is broadly consistent with the recent RICS (2002) survey. It should be noted that the forecast increase in house prices is less in the weighted model because the GTE is more expensive but faster than existing public transport and the weighted model involves increasing the sensitivity of house prices to travel cost increases (Tables 17 and 18).

4 Conclusions

This study has had a number of findings as follows:

First, the literature review confirmed that the bid choice approach that combines bid rent and choice models of residential preferences was the most appropriate way forward.

Secondly, the SP models of residential choice confirmed the importance of attributes such as transport times and costs, quality of schools, housing densities and noise levels. These models produced plausible estimates of the value of non-working travel time (in relation to travel cost) but appeared to produce implausibly high valuations of attributes in relation to house costs. We may speculate that this was due to a combination of different types of bias. Specification bias may have arisen due to omitted variables, such as those related to house type. Instrument bias may have arisen due to the unrealistic manner in which some attributes were represented in the SP experiments. Non-commitment bias may have arisen, as respondents in SP experiments are not committed to behave in the way they state in the surveys. In particular, it is difficult to represent budget and other constraints and the transaction costs associated with residential relocation in a SP experiment.

Thirdly, we found that disaggregate data did not readily exist in our Greater Oxford case study area to develop a bid choice approach at a meaningful level of spatial resolution. We are aware that house price data is becoming available at the postcode sector level (e.g. OX13 – see for example www.upmystreet.com). This in combination with 2001 Census data and improved transport models offers the prospects of more disaggregate data in the near future, which we intend to investigate in follow-up research.

Fourthly, an HP model produced more plausible valuations of the impact of travel time, quality of schools, housing densities and noise levels on house prices. However, the impact of travel costs was implausible, although not particularly statistically robust.

Fifthly, application of the HP model suggests that transport policy changes appear to have relatively modest impacts on house prices, particularly if more plausible assessments of the valuation of travel costs are used. The unweighted HP model suggested road user charging might reduce house prices on average by around 2%, although this was made up of a reduction of house prices of on average 3% outside the charged area and an increase in house prices of 2% inside the charged area. A 10% change in fuel duty was found to have a similar overall effect, leading to an average change in house prices of around 3%, but with the direction of change being uniform throughout the study area. It was also found that introducing a new public transport system (Guided Transit Express, now Expressway Oxford) might increase house prices by around 3% on average, with the greatest increases being in central Oxford and the outer suburbs (Abingdon and Kidlington) These changes may be considered modest given the backdrop of an increase in house prices over the last year of over 30%.

We therefore conclude that transport policy has a small but significant impact on the housing market. For example, we estimate that our Kidlington–Oxford–Abingdon corridor has a population of over 190,000 and almost 80,000 residential dwellings. An average price increase of £7,000 as a result of the Guided Transit Express, would suggest a windfall gain in the residential property market of over £500 million. This suggests that there may be substantial scope for fiscal measures that capture such increases in land values, though it should be noted that such measures would themselves probably affect the behavior of households and hence of the market.

Acknowledgements This work was financed by the Department for Transport's New Horizons programme. The assistance of the advisory panel (Bill Macmillan of the School of Geography and Environment, University of Oxford, Peter Headicar of Oxford Brookes University and Ian Walker of Oxfordshire County Council) is gratefully acknowledged. We are also grateful for important contributions to this project made by Francisco Martinez and David Simmonds. We also give thanks to three anonymous referees for their useful comments. All mistakes are, of course, our own.

References

Alonso W (1964) Location and land-use. Harvard University Press, Cambridge

Anas A (1982) Residential location markets and urban transportation. Academic Press, New York

Banister D, Berechman J (2000) Transport investment and economic development. UCL, London

C J Associates (2001) GTE for Oxfordshire Limited Feasibility Study. Internal Report for GTE for Oxfordshire Limited

Cervero R, Landis J (1995) Development impacts of urban transport: a US perspective. In: Banister D (ed) Transport and urban development. E&FN Spon, London

David Simmonds Consultancy in collaboration with the Transport Studies Unit, University of Oxford (2003) Impacts of urban quality on business stage 2: modelling. Technical Note 4. The impact of environmental improvements on business location decisions, TSU, University of Oxford, Oxford

Debrezion G, Pels E, Rietveld P (2006) The impact of rail transport on real estate prices: an empirical analysis of the Dutch housing market. Report, University of Amsterdam, Amsterdam

Hayashi, Y. and Doi, K. (1989) A model for analysing the imputation of consumer's benefits to land property values. In: Transport policy, management and technology towards 2001. Proceedings of the Fifth World Conference on Transport Research, vol 1, pp 303–317

Hess DB, Almeida TM (2007) Impact of proximity to light rail rapid transit on station-area property values in Buffalo New York. Urban Stud 14:1041–1068

Ingram GK (1998) Patterns of metropolitan development: what have we learned? Urban Stud 35:1019–1035

Kumar A (1980) Use of incremental form of logit models in demand analysis. Transp Res Rec 775:21–27

Martinez FJ (2000) Towards a land-use and transport interaction framework. In: Hensher DA, Button KJ (eds) Transport modelling. Pergamon, Oxford, pp 145–164

Pagliara F, Preston J (2002a) State-of-the-art technical note 1. Transport and Residential Location Project. Transport Studies Unit, University of Oxford, Oxford, January

Pagliara F, Preston J (2002b) Stated preference report technical note 2. Transport and Residential Location Project. Transport Studies Unit, University of Oxford, Oxford, April

Pagliara F, Preston J, Kim JH (2002a) Residential location choice behaviour in Oxfordshire. In: Proceedings of the European Transport Conference, Cambridge (Also Technical Note 3, Transport and Residential Location Project, July)

Pagliara F, Preston J, Kim JH (2002b) Application report technical note 5. Transport and Residential Location Project. Transport Studies Unit, University of Oxford, Oxford, December

Preston J (1991) Demand forecasting for new local rail stations and services. J Transp Econ Policy 25:183–202

RICS Policy Unit (2002) Land value and public transport summary of findings. ODPM/RICS, London

Rosen S (1974) Hedonic prices and implicit markets: product differentiation in pure competition. J Pol Econ 82:34–55

So HM, Tse RYC, Ganesan S (1996) Estimating the influence of transport on house prices: evidence from Hong Kong. Academic papers, University of Hong Kong, Hong Kong

Tinch R (2002) Ideas for a case study – Commercial values of urban quality. In TSU and David Simmonds Consultancy (eds) The effects of urban quality improvements on business performance: a literature review, expert panel and modelling study, Reference 922, Appendix C. University of Oxford, Oxford

TRL (Transport Research Laboratory) (1993) The longer term effects of the Tyne and Wear Metro. University of Newcastle, Tyne and Wear

Van de Vijvere Y, Oppewal H, Timmermans HJP (1997) The validity of hierarchical information integration choice experiments to model residential preferences and choice. Geogr Anal 30:254–272

Walmsley DA, Perrett KE (1992) The effects of rapid transit on public transport and urban development. HMSO, London

Zhang Y, Shing HF (2006) The London congestion charge and property prices: an evaluation of the impact on property prices inside and outside the zone. http://mpra.ub.uni-muenchen.de/4050/

The Influence of Accessibility on Residential Location

Jonas Eliasson

Abstract The focus of this paper is on modeling the influence of accessibility on the household's location decision. Our main theoretical contribution is an elaborate specification of what we should mean by "accessibility" in this context. This is done by assuming that households make a joint choice of location and activity pattern subject to income and time constraints. This activity pattern implies a stochastic travel pattern, the expected value of which is known at the time of location. The locational utility then consists of four parts: the indirect utility of income and time net of housing cost and expected total travel time and travel cost, the direct utility of the optimal activity pattern, the direct disutility of the expected travel pattern and the direct utility derived from location characteristics. The locational utility is then used in a discrete choice model for the choice of location.

In the empirical part of the paper, we present methodology and results from the estimation of TILT, Tool for Integrated analysis of Location and Travel, a land use-transportation model for the Stockholm region. Among other things, we find that the attractiveness of a location increases both with the accessibility to workplaces and with the accessibility to different types of service.

1 Introduction

The subject of this paper is modeling the influence of accessibility on a household's location decision. Our primary theoretical contribution is an elaborate specification of what we should mean by "accessibility" in this context. We propose a framework that integrates a stochastic travel pattern of the discrete choice type in an optimization model for the joint choice of location and optimal activity pattern. Further, we show how a non-linear indirect utility function can be introduced in a way that is

J. Eliasson
Centre for Transport Studies, Royal Institute of Technology, Stockholm, Sweden
e-mail: jonas.eliasson@abe.kth.se

F. Pagliara et al. (eds.), *Residential Location Choice*, Advances in Spatial Science,
DOI 10.1007/978-3-642-12788-5_7,

consistent with all the choices involved. On the empirical side, we present methodology and results from the estimation of an instance of the framework, a land use-transportation model for the Stockholm region called TILT, Tool for Integrated analysis of Location and Travel. Apart from being an example of how the proposed framework can be made operational, we show that accessibility to workplaces as well as accessibility to service, shopping etc. are all important determinants of the attractiveness of a location. We also suggest a measure of the preference for the number of rooms relative to the floorspace.

The theoretical framework of TILT has previously been presented in Eliasson and Mattsson (2000), together with a number of numerical simulations comparing it to a standard nested logit model. Although the framework is almost exactly the same in this paper, the derivation and interpretation of it is different on a number of points. An earlier version of the framework was presented in Eliasson and Mattsson (1998). Methodology and results from the estimation of the travel model of TILT are presented in Eliasson (2000).

In the present paper, we will focus on modeling the location choice of households, and in particular the connection between travel and location. The outline of this paper is as follows. In Sect. 1, we place the paper in a context of earlier contributions in the field of location and transportation modeling. In Sect. 2, we propose a framework for a household's joint choice of location and activity/travel pattern. Section 3 summarizes the estimation of the travel model that underlies the location model. In Sect. 4 we report and discuss estimation methodology and results. Section 5 concludes.

1.1 Modeling Household Behavior

Before we begin presenting the model, let us consider for a moment the complexity of a household's decision. This will serve as a background, and as a motivation for the necessity to limit the scope of a model.

A household's decisions of residential location, workplace, activities and travel pattern are a inextricably entangled weave of mutual interdependencies and constraints. Each of these choices is connected to all the others, and each one consists of not one single choice but a range of options, all depending on each other and with varying degrees of similarity and substitutability. Moreover, the choices are subject to a multitude of constraints, such as budget constraints on long and short term, time constraints and various scheduling constraints.

Depending on what time scale we consider, the nature of these constraints and interdependencies changes, and it is not even obvious what can be "chosen" and what is "given". Is car ownership given or chosen? In the short run, it is given; in the long run, it is chosen.

If we consider the decisions made during a single day, an individual may have to be at work at a fixed time in the morning and to pick up children in the afternoon. Constrained by these activities, and hence subject to a *time budget* constraint other activities have to be scheduled: shopping, cooking, eating, sleeping and so on. Choosing the location and duration of one of these activities depends on the location and duration of all the others. There are also dependencies between the activities. One activity may be necessary in order to carry out another. Two types of activities may be related in the sense that they are more or less close substitutes for each other. Most of these choices are also subject to household interactions and negotiations: Who picks up the children? Who does the shopping?

The choice of activity pattern is made even more complicated by the need to construct a feasible travel pattern to accommodate these activities. The trips needed to carry out the activities can then be more or less desirable in themselves, consisting of anything from waiting in the rain for a bus to taking a walk in the park.

There is also a *monetary budget* constraint constraining all of these choices, although it is difficult to specify exactly how. Over the course of a longer time-period – a year or maybe several years – expenses, savings and income have to be equal. It is not obvious how this long-term constraint should be translated into something that applies to a shorter time period – a single day or a week.

Turning to the long time scale, maybe several years, the household may choose another residential location, workplace and school, whichever applies. The household itself may change – children are born, grow up and leave home, couples get married and divorced. The budget constraint is even worse to specify than before, since it is uncertain what the household's income and expenses will be in the years to come. When comparing residential locations, the household has to evaluate the attractiveness of the location by solving an activity pattern problem of the type we described above conditional on this location, taking into account the constraints caused by both the geographical location and the housing expenses. Moreover, this predicted activity pattern (and the travel pattern associated with it) will not be known precisely at the time of location. The household has only limited knowledge of its future preferences, its precise activity and travel alternatives, and its time, budget and scheduling constraints.

Furthermore, most of the choices we have mentioned above are also dependent on the choices of other households, either as explicit market processes (e.g. the housing market) or through externalities (e.g. traffic congestion and location externalities).

The task of an urban modeler is therefore clearly so difficult that it is almost ridiculous. What we want to do is to replicate not only *one* household's behavior in terms of residential choice, travel pattern and so on; we want to replicate the interdependent choices of maybe *millions* of households. And more than that, we want to predict not only what they do *today*, but what they would do under *different circumstances*: if their income change, if travel times or travel costs change, if a new shopping mall opens, if the number of working hours a day changes, if they get the possibility to telecommute and so on.

This is not "almost" ridiculous – it's absurd. Nevertheless, we have no choice but to try. We need information about the consequences of various planning decisions, decisions ranging from bus timetables and intersection signals to zoning regulations and road construction. We need information of a type that goes beyond intuition, hopes and guesses. Modeling the behavior of the households and individuals is the best way – and sometimes the only way – we have to get this kind of impact assessment.

1.2 The Scope of a Model

This background serves as a motivation for two somewhat contradictory reflections. First, it is necessary to limit the scope of a model. Models with very large scopes, trying to capture everything from trip scheduling to household formation in one framework, will run into problems with data availability, validation, estimation and theoretical and empirical consistency. Second, confining the scope of a model too much, keeping related choices fixed while focusing on some conditional decision, will tend to underestimate the effects of a change. Think of a model that predicts route choice while keeping mode and destination choice fixed. If we use such a model to predict consequences of a large-scale change in the transportation system, the impacts of these changes will clearly be underestimated. It also increases the risk of specifying models that have no clear behavioral underpinning. This is sometimes called the "statistical fallacy": to include just about any variable in an econometric estimation as long as it results in significant parameters. Think of estimating the probability that an individual will choose a particular residential location. Clearly, there are many variables affecting this choice, and they can be included in the estimation in a large number of ways. Some of these ways will be "consistent", in the sense that they are compatible with an underlying behavioral model, and some will not. How we include the explanatory variables will in general have a considerable impact on the model's predictions.

The scope of a model should be adjusted according to the questions we want to analyze. For some questions, it may be necessary to try to capture the precise scheduling and duration of a household's activity pattern, while it is sufficient to treat long-term choices such as residential location and workplace as fixed. For other questions, choices like activity scheduling and route choice may be of secondary importance, while the emphasis should be on capturing long-term processes like household formation and residential relocation.

In this perspective, it is natural to classify different models according to which choices they keep fixed. For example, the traditional approach of urban economics is to treat the travel pattern as fixed (often just as a daily car trip to the city center), while studying the choice of residential location. Conversely, the traditional approach of transport models is to keep residential location fixed, while studying the choice of destination, mode and travel route, or often just one or two of these choices.

Another important classification of models is the way they treat the dynamics in the system. Some models explicitly model the dynamics, often through a simulation approach. Other ones model an equilibrium situation to which the system is assumed to converge, for example traffic network equilibrium models or supply-demand equilibrium models. Equilibrium models can often be embedded in a dynamic framework using discrete time steps. The system is then assumed to converge to a short-run equilibrium in each time step, while there are long-term processes changing over time. This approach has for example been used to model housing markets, where the prices are assumed to be set to make short-run demand equal short-run supply, while housing stocks change at a much slower rate, driven by the current market price for housing.

1.3 The Context and Purpose of the Paper

With this perspective as a background, we will now describe the purpose of the present paper, placing it in a context of more or less similar contributions. As our focus is on the household's location choice, we will not consider the housing supply side, assuming that housing prices and characteristics are given. On the travel side, we will limit the scope of the model to the choice of destinations, frequencies and modes for a number of trip types, excluding route choice and trip scheduling. Excluding trip scheduling means that we will not explicitly model durations and timing of activities, although it is easily shown that our model is consistent with a hypothetical underlying model including such considerations.

Reviews of the land use-transportation modeling literature are in ample supply, for example Hunt et al. (2005), Chang (2006) and Wegener (1994a, b, 1998, 2004). Here, we will only make a brief summary of how the way accessibility is assumed to influence the attractiveness of a location has developed over time. The mathematics will be sketchy, merely clarifying the points we wish to make. Formal and comprehensive treatments of the material presented here can for example be found in Fujita (1989) (urban economic theory), Anas (1982) (discrete choice location models), and Ben-Akiva and Lerman (1985) (logit models).

What should we mean by the "accessibility" of a location? A fruitful way to sort things out is to distinguish between *activity pattern* and *travel pattern*.[1] An activity pattern is a vector of activities, like work and shopping, each of which is carried out at some particular location between two points in time. This activity pattern will then be associated with a travel pattern, a vector of trips between points in space-time with some travel mode and along some route. It is the activity that is the source of utility, i.e. the reason to make a trip at all, either directly or indirectly. For example, a trip may be caused by the wish to carry out an activity that generates utility, or to buy something that will generate utility. The trip will be associated with a *generalized cost*, consisting of a resource cost, which is the time and cost devoted to the trip, and a direct disutility. The former stems from the fact that the trip will

[1]This approach draws on the work by Becker (1965), De Serpa (1971) and Evans (1972). The "framework" sketched here, though, keeps just the parts of their work that we will need in this context. One of the main omissions is the interdependencies between different types of activities, where one activity can be a prerequisite of another.

decrease the resources (time and money) that can be spent on other consumption and activities. The latter is associated with characteristics of the trip such as waiting times, number of changes, car queues and so on, and reflects for example that waiting time at a bus stop is an inconvenience apart from just the time that is spent on waiting (which is captured in the resource cost). In some cases, there can be a direct utility associated with a trip, rather than a disutility; for example, there are reports of people who enjoy walking. The generalized cost will depend on travel times, travel costs etc., but also on available income and time and total travel costs and travel times, which will determine the marginal utilities of time and money.

It is a crucial question how the activity pattern is assumed to generate the travel pattern. Obviously, there are in general many travel patterns that are compatible with a given activity pattern. But which of these travel patterns will be chosen? The standard assumption, which we will use here, is that the travel pattern with the least generalized cost will be chosen. We will say that the least-cost travel pattern compatible with a given activity pattern is the travel pattern *implied* by the activity pattern.

We can formalize this by letting $u_1(y,t)$ be the indirect utility of spending income y and time t on goods and activities, $u_2(\mathbf{s})$ be the utility of an activity pattern $\mathbf{s}$, and p_i the housing cost at location i. Let $\mathbf{b}_i$, $\mathbf{t}_i$ and $\mathbf{q}_i$ be vectors of the costs, times and direct disutilities of the possible trips that can be made from location i, and $\mathbf{x}$ the travel pattern implied by $\mathbf{s}$. The "accessibility" of a location i can then be defined to be the indirect utility of residual income (available income minus housing costs and total travel costs) and residual time (available time minus total travel times) plus the utility derived from the optimal activity pattern minus the direct disutility of trips[2]:

$$u_i = u_1(Y - p_i - \mathbf{b}_i\mathbf{x}, T - \mathbf{t}_i\mathbf{x}) + u_2(\mathbf{s}) - \mathbf{q}_i\mathbf{x}. \tag{1}$$

$\mathbf{b}_i\mathbf{x}$ is the scalar product of travel costs and trips, giving the total travel cost for the travel pattern. Similarly, $\mathbf{t}_i\mathbf{x}$ and $\mathbf{q}_i\mathbf{x}$ are the total travel time and total direct disutility. In many of the contributions we review here, the travel pattern will be fixed; the activity pattern will almost always be. If they are not, the approach we described above can be formalized in the following way. Let $\boldsymbol{\xi}(\mathbf{s})$ be the set of travel patterns compatible with the activity pattern $\mathbf{s}$.

$$u_i = \max_{\mathbf{s}} \max_{\boldsymbol{\xi} \in \boldsymbol{\xi}(\mathbf{s})} \{u_1(Y - p_i - \mathbf{b}_i\boldsymbol{\xi}, T - \mathbf{t}_i\boldsymbol{\xi}) + u_2(\mathbf{s}) - \mathbf{q}_i\boldsymbol{\xi}\}. \tag{2}$$

This is a two-step optimization: the optimal travel pattern $\mathbf{x} = \boldsymbol{\xi}^*$ is chosen conditional on the activity pattern $\mathbf{s}$, and the activity pattern is chosen to maximize the locational utility u_i.

[2]The direct utility of the location is omitted here, and with that the influence of other location characteristics than accessibility. We will for example not treat the possibility to endogenously determine lot sizes, which was one of the main interests of early urban economics. Neither will we treat the implications of freely chosen working hours; these are considered to be fixed here.

We now turn to how earlier studies fit into this framework. The dedicated study of the interactions between household location, accessibility and land price began in the 1960s with the contributions by Alonso (1964), Mills (1967) and Muth (1969). At this early stage, the "activity pattern" was simply to be at work each day and at home each night, and the travel pattern was merely a daily commuting trip by car to the workplace in the city center. In the notation above, **s** (and thus $u_2(\mathbf{s})$) was fixed, and **q** was set to zero, leaving only the first term. (In fact, it was first with Evans (1973) that residual time was included in the indirect utility function.)

This simple "travel pattern" has later been extended, first to several travel modes, then to several destinations and then to a general travel pattern (e.g. Jara-Díaz and Martínez 1999). In these extensions, the travel pattern is to some extent different depending on location. The focus is still exclusively on the first term of (1), however; it is left unexplained why the household makes any trips at all, since travel only means less money and time left for other things.

Most of this tradition was purely theoretical, providing insights rather than forecasts, and founded on general observations rather than dedicated investigations. The 1960s also saw the first attempts at operational models, going beyond comparative statistics and pencil-and-paper solutions. These attempts were only mildly successful, however, both from a theoretical and an empirical point of view. The introduction chapter in Anas (1982) provides an interesting account of the early stages of the relationship between the theoretical urban economics school and the empirically oriented urban modelers.

Anas' book was also an early application of a new way to model the location choice, advocating the use of *discrete choice* models based on *random utility* (McFadden 1974). This introduced measurement of accessibility through the *expected generalized* travel cost. The early random utility models and the urban economics models still had in common that the demand for travel was not modeled endogenously, but with the much more econometrically oriented random utility models, the third term of u_i in (1) – the direct disutility of the trip – was considered.

The expected generalized cost approach can be derived by assuming a linear indirect utility function $u_1(y,t) = \lambda y + \sigma t$. Dropping terms independent of i (since they cancel out when comparing different locations), the locational utility becomes

$$u_i = -\lambda p_i - (\lambda \mathbf{b} + \sigma \mathbf{t} + \mathbf{q})\mathbf{x}. \tag{3}$$

$\lambda\mathbf{b}+\sigma\mathbf{t}+\mathbf{q}$ is the vector of *generalized costs*. The contribution of the random utility models was now that this generalized cost was supposed to be *stochastic*. The assumption is that in addition to the measurable parts of the generalized cost – travel times, travel costs etc. – there is a random term, unobservable to the modeler. The accessibility was then measured by the expected generalized cost, which if the random terms of the generalized cost is assumed to be independently Gumbel distributed becomes the familiar logsum. This way, a location choice model was obtained that was consistent with the successful travel models using discrete choice models.

There is a caveat here, though. The expected generalized cost approach depends crucially on the assumption of a linear indirect utility function. It is not evident how this approach can be extended to a nonlinear indirect utility function. This pitfall seems to have received less attention than it should. For example, consider a model of this type:

$$u_i = u_1(Y - p_i) - (\lambda \mathbf{b} + \sigma \mathbf{t} + \mathbf{q})\mathbf{x}. \tag{4}$$

This locational utility consists of a non-linear function of income minus housing expenses, minus the expected generalized cost (a logsum measure, for example). This will in general *not* be consistent, since there is no guarantee that the marginal indirect utility of money will be the same throughout the model. Another problem with a nonlinear indirect utility function is how the stochasticity in the underlying travel model will carry over to the locational utility function. If the travel cost is stochastic, the travel pattern will be stochastic, and so will the residual income and time. It is not evident how this stochasticity will carry over to the locational utility function. A third problem is if we want to consider a travel pattern consisting of several trips. If the indirect utility function is nonlinear, then the marginal utility of money will be non-constant. This means that the various travel choices will affect each other, since they share the same marginal utilities of time and money, and these marginal utilities will depend on total travel costs and times.

Another issue that seems to have received relatively little attention is the relationship between the activity pattern and the travel pattern. In almost all applications, the activity pattern is fixed to be for example five work trips a week, and perhaps a number of shopping trips. In fact, the most common case is that the underlying assumptions of a travel pattern are not explicit. Expressed in the terms of the locational utility (1), the term $u_2(\mathbf{s})$ has been neglected.

The main theoretical contribution of this paper is an investigation of all these issues. We propose a model in which each household makes a joint choice of location and activity pattern, subject to time and budget constraints. This activity pattern will then imply a stochastic travel pattern, which is not known deterministically by the individual. We will see how this approach solves the problems indicated above.

In the past few years, there has been a growing interest in obtaining models with a more direct connection between the choice of location and the choice of a travel pattern, and modeling several trip types in an integrated way. An early and influential model of this type is NYSIM (Anas 1995), which has been applied in several studies. In NYSIM, location choice is made conditional on workplace location. The utility of a location depends on expected generalized travel costs for work trips and the utility derived from the optimal shopping trip pattern. The theoretical framework of Jara-Díaz and Martínez (1999) also fits into this line, although they do not explicitly consider how the spatial distribution of the optimal activity pattern is obtained. Their paper also fits into another recent trend, which we also have argued for above, the growing interest in viewing travel as a derived demand. Rather than modeling trips as a “good”, from which the household derives its utility directly, trips are viewed as necessary in order to perform other, utility-generating (or at least income-generating) activities. That travel should be viewed as a derived demand is also one of the two main motivations for the development of *activity-based* travel models (see Ettema and Timmermans 1997; Axhausen and Gärling 1992; Algers et al. 2005), the other being the striving to incorporate time-geographic and scheduling constraints into the travel pattern. Ben-Akiva and Bowman (1998) combine a location choice model with an activity-based travel model.

1.4 Price Determination and Location Choice

Before we start to develop the model framework, however, we need to say something about how we assume that households choose between locations and how we assume that housing prices are determined. There are primarily three major ways to do this.

The first way is the *discrete choice* approach, pioneered by McFadden (1978), Anas and Moses (1979) and Anas (1982). It has since been used in many operational models, either as models for joint choice of location, workplace and mode choice (e.g. Abraham and Hunt 1997), or as an "add-on module" where households choose location and dwelling type according to a discree choice model, but the transport-related choices are handled through some sort of accessibility measure rather than as a joint choice (e.g. Jonsson 2008; Waddell 2002). Here, the households calculate the locational utility of each possible location, and choose the one yielding maximum utility. The households view the prices of locations as given, just as in standard neoclassical economics. The modeler can only observe the locational utilities up to a constant, which varies across households. The locational utility is thus assumed to be $U_i = u_i + \varepsilon_i$, where u_i is the locational utility measured by the modeler, called *strict* utility, and ε_i is modeled as a random term, known to the individual but not to the modeler. The modeler can thus only predict the location choice probabilities $P^n{}_i(\mathbf{p})$, the probability that household n chooses location i given the price vector $\mathbf{p}$. The expected aggregated demand becomes $D_i(\mathbf{p}) = \Sigma_n P^n{}_i(\mathbf{p})$. The price vector is then calculated to make demand equal supply (which can also depend on $\mathbf{p}$). If locations consist of a fixed number of individual dwellings (rather than zones or some other aggregate), then $\Sigma_n P^n{}_i(\mathbf{p}) = 1$ for each i. If households are identical, this implies that prices are determined such that the strict locational utility $u_i(\mathbf{p})$ is equal for all i.

Urban economics, on the other hand, also assume that households choose their utility-maximizing location. But instead of working directly with the locational utility u_i, it is inverted to calculate a *bid rent* $r(u,i)$, defined to be the maximum price a household could pay for location i and still achieve utility u. Location i is sold to the highest bidder. Market prices are determined by the equilibrium condition that no household should be able to be better off by changing location. If households are identical, this implies that prices are determined such that the locational utility $u_i(\mathbf{p})$ is equal for all i, just as in the discrete choice approach.

This approach appeared first as a theoretical construct in the theoretical urban economics tradition, preceding the discrete choice approach. The approach was then applied in an operational model by Ellickson (1981), and then in a stochastic formulation by Martínez (1996). Martínez (1992) showed the consistency between the discrete choice approach and the bid-rent approach under equilibrium assumptions. A similar approach has been used by Waddell (2000, 2002).

In this paper, we will focus exclusively on the demand side, i.e. the household's choice of location. We will assume that households take prices as given. However, the way prices are formed on a competitive land market has implications on the econometrics. The crucial difference between location choice models and for example mode or destination choice models is that each of the alternatives is actually chosen by some household (in equilibrium). This is a significant difference from, say, mode choice models, where we are able to include alternatives chosen by nobody by generating them from travel supply data. In fact, were it not for differences across households such as income and family characteristics, the observable part of the utility u_i would be the same for all residences, assuming market equilibrium prices. Generalized travel costs and environmental characteristics will be capitalized into housing prices. Thus, it is evident that the explaining power of the model is largely determined by how finely described the households are, as opposed to, say, mode choice models, where the travel time and travel cost of an alternative is often able to explain the observed choices to a large extent.

In passing, we note that the observation that prices are set such that u_i is equal everywhere (if households are identical) is the cornerstone of *hedonic price* studies, where the price is explained as a function of the locational attributes. We will not go into this line of research here; see Rosen (1974) and Wheaton (1977) for its foundations.

The third way to model the housing market is *microsimulation*. Here, there is no assumption of market equilibrium. Instead, the relocation process of individual households is modeled at the microlevel, as a decision sequence typically consisting of choosing whether to move, bidding on a new location, and a simulated bargaining process on the price with the seller. Locations and prices are then the total outcome of all these microlevel simulations. Thus, there is no explicit equilibrium present, only a large number of individual contracts.

2 An Integrated Framework for Location and Travel Choices

The basic assumption of the model is that, when choosing location, each household makes a joint choice of location and activity pattern. This activity pattern implies a stochastic travel pattern. At the time of the location choice, only the expected travel pattern is known. The choice is subject to time and budget constraints. There are two major advantages with treating all of these choices in one coherent framework compared to treating location and travel decisions as a series of isolated choices.

The first is that this allows the marginal utilities of time, money and different kinds of trips to depend on the entire activity and travel patterns. This ensures consistency between the different choices, and also introduces dependencies between the various choices. For example, increasing the number of leisure activities at a certain destination, and hence the number of leisure trips there, might influence other travel and activity choices in several ways. There is less time and money left to spend on other things, including travel, so the marginal utilities of time and money will increase. The marginal utility of leisure activities might decrease, and we might expect that the marginal utility of leisure activities at this particular destination would decrease more than the marginal utility of leisure activities in general.

The second advantage is that this introduces an explicit connection between location and trip generation. It is fairly common that location models introduce some "accessibility measure" directly into the indirect utility function, but without specifying *what travel pattern a certain accessibility will generate*. In other words, the correspondence between the underlying travel pattern and the accessibility measure is unclear. Here, we will let the utility of a location depend directly on the utility of the optimal activity pattern, the indirect utility of residual income and time and the direct disutilities of the expected travel pattern. This establishes an explicit connection between the choice of location and the choice of activity and travel pattern.

For the moment, we do not distinguish between different household segments to reduce the number of indices. Naturally, many of the functions introduced below can be different across household segments, where the segmentation can be according to number of children, number of adults, employment status etc. Throughout the paper, we will not consider household interactions, but assume that we can model the household with a representative individual. The number of adults in the household will, however, influence preferences for type of dwelling (Sect. 4). (Ideally we should also distinguish between one- and two-worker households, but our data does not contain information about the employment status of spouses.) (The following derivation of the model follows closely that in Eliasson 2000.)

Call the individual's available income per time period Y and its available time per time period T. Minimal living costs (taken from the Swedish Consumer Agency 1997), are subtracted from available income Y. For households with two adults, Y is half of the household income. Working hours are assumed to be fixed, so Y includes working income, and working hours are subtracted from T – hence, T will be different for non-workers, part-time workers and full-time workers. y and t are the residual time and residual income, the time and money devoted to activities and consumption other than travel and housing expenses. Let $u_1(y,t)$ be the indirect utility function, i.e. the utility of spending y and t.

Our choice alternatives will be individual dwellings rather than zones or house types. Although accessibility data are generally given on some more aggregate (zonal) level, we have data on characteristics of the individual dwellings, such as floorspace, price and number of rooms. Let $\mathbf{K}_i$ be a vector of quality characteristics describing residence i, such as environmental variables, number of rooms and so on. Let $u_3(\mathbf{K}_i)$ be the utility derived from these residence characteristics, and p_i the residential cost (rent or equivalent) per time period.

The activity pattern consists of frequencies and locations for each activity $k = 1\ldots\kappa$. For example, the activity types we will use in the subsequent application is work, school, shopping, service, leisure, social and "other". This categorization can certainly be refined if need be. For each activity type, we will distinguish between two sub-types, *location-generic* and *location-specific* activities, and assume that a certain fraction of the k-activities is location-generic, while the rest of the k-activities are location-specific.

Intuitively, an activity is location-generic if it does not matter where it is carried out. Formally, this means that the utility of performing an activity depends only on the *total* performed "amount" of this activity, but not on *where* the activity has been undertaken. For example: Having made one, say, grocery shopping trip means that an additional shopping trip might seem less attractive. We might expect, though, that the attractiveness of a particular destination is virtually independent of the number of previous shopping trips to that destination.

For location-specific activities, on the other hand, the marginal utility of performing an activity does depend on where the previous activities have been carried out. For some activity types – we can think of recreational travel or non-daily shopping – there is an *incentive for variation*. Usually, people do not want to visit the same friend, shoe shop or cinema over and over again.

Let z^k be the number of location-generic k-activities, $s^k{}_j$ the number of location-specific k-activities in zone j, and $u_2(\mathbf{s},\mathbf{z})$ the utility of the activity pattern $\mathbf{s} = \{s^k{}_j\}$ and $\mathbf{z} = \{z^k\}$. The total number of k-activities is then $z^k + \Sigma_j\, s^k{}_j$.[3]

The activity pattern ($\mathbf{s}$,$\mathbf{z}$) implies some travel pattern $\mathbf{X} = \{X^k{}_{ljm}\}$, consisting of trips from zone l to zone j in order to perform activity k. We will describe the mapping from activities to trips with a vector-valued function $\mathbf{\Phi}^i(\mathbf{s},\mathbf{z}|y,t,\mathbf{B},\mathbf{T},\mathbf{Q})$. This mapping depends on residential location i, the residual income and time y and t, and the vectors of travel costs, times and disutilities $\mathbf{B} = \{B_{ijm}\}$, $\mathbf{T} = \{T_{ijm}\}$ and $\mathbf{Q} = \{Q^k{}_{ijm}\}$. In its simplest form, $\mathbf{\Phi}^i$ can just be one trip per activity, from home to where the activity is performed; this is the approach we will use here. In more

[3]Note that we will only model the number and location of activities, not their durations or scheduling in time. It could very well be argued that for each activity, an amount of time and money must be spent. In the simplest case, we could introduce constraints requiring that each time an activity k is carried out, some amount of money πk and time νk must be spent. Such constraints (or more complicated ones) are called technical constraints, and are sometimes useful in theoretical frameworks (see e.g. De Serpa 1971, or Evans 1973). However, we seldom or never have information of the nature of these constraints, and they are almost always very flexible, in the sense that the money and time spent on an activity can be chosen rather freely. Instead, we assume that the residual income and time y and t are distributed optimally over activities and goods, and the resulting utility of this is u1(y,t) + u2(s,z). The dubious assumption is then that these function are separable. However, the analysis conducted here can easily be extended to the case of inseparable utility functions.

elaborate travel models, we can introduce various forms of trip chaining, all the way up to a full-fledged simulation model, including various time-geographic constraints and so on.

B_{ijm} and T_{ijm} are the travel cost and travel time from i to j with mode m. $Q^k{}_{ijm}$ is the direct disutility of a k-trip with mode m between i and j. The $Q^k{}_{ijm}$:s are functions of trip characteristics like waiting times, inconveniences like riding a bicycle on a shopping trip, and socioeconomic characteristics like sex and license holding. Travel costs, travel times and travel utilities are all assumed to be stochastic, in the sense that they vary a little from trip to trip, and are not completely known neither to the household nor individuals at the time of the location choice. We will let uppercase letters denote stochastic variables and lowercase letters their expected values, i.e. $b_{ijm} = \mathrm{E}(B_{ijm})$, $t_{ijm} = \mathrm{E}(T_{ijm})$, $q^k{}_{ijm} = \mathrm{E}(Q^k{}_{ijm})$.

This means that while the activity pattern is completely determined at the time of decision, the travel pattern is not. Instead, the travel pattern $\mathbf{X}$ is *stochastic* from the individual's point of view, and the function $\boldsymbol{\Phi}^i(\mathbf{s},\mathbf{z}|y,t,\mathbf{B},\mathbf{T},\mathbf{Q})$ is a stochastic function (in the sense that it depends on the stochastic variables $\mathbf{B}$, $\mathbf{T}$ and $\mathbf{Q}$). This is because the household is assumed to have limited information about future trip costs, times and direct utilities. Let $\mathbf{x}$ be the expected travel pattern $\mathrm{E}(\mathbf{X})$. We will get back to the distributions of these variables in the next section.

This is the activity-travel optimization problem conditional on residence i:

$$\max_{y,t,\mathbf{s},\mathbf{z}} \left[u_1(y,t) + u_2(\mathbf{s},\mathbf{z}) - E\left(\sum_{kljm} X^k_{ljm} Q^k_{ljm} \right) + u_3(\mathbf{K}_i) \right] \text{s.t.} \tag{5}$$

$$y + p_i + E\left(\sum_{kljm} X^k_{ljm} B_{ljm} \right) = Y \tag{6}$$

$$t + E\left(\sum_{kljm} X^k_{ljm} T_{ljm} \right) = T \tag{7}$$

$$z^k \geqslant 0 \ \forall \ k \tag{8}$$

$$s^k{}_j \geqslant 0 \ \forall \ j,k \tag{9}$$

where by definition

$$\mathbf{X} = \Phi^i(\mathbf{s},\mathbf{z}|y,t,\mathbf{B},\mathbf{T},\mathbf{Q}). \tag{10}$$

The constraints (8)–(9) can of course be replaced by more general maximum/minimum constraints, if we have data on such constraints. Especially for work trips, it can be natural to assume that the trip frequency is constrained to a fixed number, either on the form $s^{work}{}_j = constant$ if the workplace location is known in advance, or on the form $z^{work} = constant$ if it is not.

2.1 Specifying the Activity-Trip Mapping

We will now describe how the travel pattern **X** depends on the activity pattern (**s**,**z**), i.e. the mapping $\mathbf{X} = \mathbf{\Phi}^i(\mathbf{s},\mathbf{z}|y,t,\mathbf{B},\mathbf{T},\mathbf{Q})$. Assume that the individual makes one home-based trip per activity. The crucial assumption is that the choices of mode in the case of trips to location-specific activities **s** and the choices of mode and destination in the case of trips to location-generic activities **z** are postponed until the time the trip is about to be made. When the trip is about to be made, the mode (and destination, in the case of location-generic trips) with the least generalized cost $C_{ijm}{}^k$ is chosen. Just as the costs, times and direct travel utilities, the generalized cost is a stochastic variable, redrawn for each trip the household makes. The generalized cost for a k-trip between i and j with mode m is defined to be

$$C_{ijm}^k = \lambda B_{ijm} + \sigma T_{ijm} + Q_{ijm}^k. \tag{11}$$

λ is the marginal utility of money and σ the marginal utility of time. Note that these are not constant, unless $u_1(y,t)$ is a linear function.

For location-specific trips (i.e. trips going to location-specific activities), the trip maker chooses the mode m with the least generalized cost $C_{ijm}{}^k$ given origin i, destination j and trip purpose k. For location-generic trips, the trip maker chooses the mode m and destination j with the least generalized cost $C_{ijm}{}^k$ given origin i and trip purpose k. Since the generalized cost is stochastic, the travel pattern $X_{ijm}{}^k$ will also be stochastic. Once we specify the distribution of the generalized costs $\{C_{ijm}{}^k\}$, we can (in principle) obtain choice probabilities $P_{m|ji}{}^k$ and $P_{mj|i}{}^k$. With these, we can write the expected travel pattern as

$$x_{ijm}^k = E\left(X_{ijm}^k\right) = z^k P_{jm|i}^k + s_j^k P_{m|ij}^k. \tag{12}$$

We will let $\mathbf{x} = \{x_{ijm}{}^k\}$ be the expected travel pattern, while $\mathbf{X} = \{X_{ijm}{}^k\}$ is the underlying stochastic travel pattern.

The generalized cost can be separated into two components. The first two terms of (11) constitute the resource cost of the trip. This cost is the decrease in utility caused by the decrease in time and money available for other consumption and activities. The less residual amount of time and money the individual or household has, the higher the marginal utilities λ and σ be, and the higher will this cost be perceived. Little residual income can either be caused by low income or by high expected travel expenses, or a combination of both. Similarly, little residual time can either be caused by long working hours or by high expected total travel times, or a combination.

The second type of the cost, the third term of (11), is the direct disutility of a trip, reflecting that the trip may be a utility or disutility in itself, apart from the time and money spent.

Now, we assume that the distributions of the costs, times and direct utilities are such that the generalized cost for location-specific trips can be written as

$$C^k_{ijm} = c^k_{ijm} + \varepsilon_m, \tag{13}$$

where ε_m is a negatively Gumbel distributed stochastic variable[4] with $E(\varepsilon_m) = 0$, and

$$c^k_{ijm} = E\left(C^k_{ijm}\right) = \lambda b_{ijm} + \sigma t_{ijm} + q^k_{ijm}. \tag{14}$$

$c_{ijm}{}^k$ is called the strict generalized cost, and is the part of the $C_{ijm}{}^k$ that the modeler can measure and the individual can predict when choosing location. The random terms ε_m are assumed to be unobservable to the modeler and unknown to the individual at the time of the location choice. Once a (k,i,j)-trip is about to be made, a vector of stochastic terms $\{\varepsilon_m\}$ is drawn.

For location-generic trips, we assume in a similar way that

$$C^k_{ijm} = c^k_{ijm} + \varepsilon_{jm}, \tag{15}$$

where ε_{jm} is negatively generalized extreme value distributed with $E(\varepsilon_{jm}) = 0$ and $c_{ijm}{}^k$ was defined in (10).

These choices of distributions mean that we can calculate the choice probabilities in (8). $P_{m|ji}{}^k$ becomes a logit choice probability, and $P_{mj|i}{}^k$ a nested logit choice probability[5]:

$$P^k_{mj|i} = \frac{e^{\mu^k_2 W^k_m}}{\sum\limits_m e^{\mu^k_2 W^k_m}} \frac{w^k_j e^{-\mu^k_1 c^k_{ijm}}}{\sum\limits_j w^k_j e^{-\mu^k_1 c^k_{ijm}}} \tag{16}$$

$$P^k_{m|ji} = \frac{e^{-\mu^k_3 c^k_{ijm}}}{\sum\limits_m e^{-\mu^k_3 c^k_{ijm}}}, \tag{17}$$

where

$$W^k_m = \frac{1}{\mu^k_1} \ln \sum_j w^k_j e^{-\mu^k_1 c^k_{ijm}}. \tag{18}$$

---

[4]When we say that ε is negatively Gumbel distributed, we mean that (−ε) is Gumbel distributed. The same convention is used for the GEV distribution.

[5]We assume that the generating function in the GEV distribution is chosen such that the nested logit choice probability is obtained. Details can be found in McFadden (1978). For presentational purposes, we assume that the nested model has mode choice at the upper level. The structure is of course determined during estimation. In the present application, mode choice happens to be at the upper level.

w_j^k is a measure of the relative size of zone j with respect to k-trips. This corrects for the bias that would otherwise be introduced by the fact that each alternative destination zone j is in fact an aggregate of several "elemental" destinations, and that the number of such elemental destinations is different across zones. They are normalized such that $\Sigma_j w_j^k = 1$. We will describe these measures later on.

The expected generalized cost for a location-specific trip becomes

$$c_{ij}^k = E\left[\min_m \left\{C_{ijm}^k\right\}\right] = -\frac{1}{\mu_3^k} \ln \sum_m e^{-\mu_3^k c_{ijm}^k}. \tag{19}$$

The expected generalized cost for a location-generic trip becomes

$$c_i^k = E\left[\min_{j,m} \left\{C_{ijm}^k\right\}\right] = -\frac{1}{\mu_2^k} \ln \sum_m e^{\mu_2^k W_m^k}. \tag{20}$$

It can also be shown that

$$E\left(X_{ijm}^k C_{ijm}^k\right) = \sum_k z^k c_i^k + \sum_{jk} s_j^k c_{ij}^k. \tag{21}$$

Proving this is straightforward but a bit lengthy. A proof can be found in Eliasson and Mattsson (2000).

2.2 Solving the Activity-Travel Problem

We will now derive the solution and the optimal value of the activity-travel problem. The optimal value u_i will be the maximal utility that the household can achieve if choosing location i. The solution will consist of the optimal activity pattern ($\mathbf{s}$,$\mathbf{z}$), its expected travel pattern $\mathbf{x}$ and the money and time (y,t) devoted to other things than travel.

Assume that all activity types are essential, so we can drop the nonnegativity constraints (8)–(9). The Lagrangian becomes

$$\begin{aligned} L = u_1(y,t) + u_2(s,z) + E\left(\sum_{jkm} Q_{ijm}^k X_{ijm}^k\right) \\ -\lambda\left\{y + p_i + E\left(\sum_{jkm} B_{ijm} X_{ijm}^k\right) - Y\right\} \\ -\sigma\left\{t + E\left(\sum_{jkm} T_{ijm} X_{ijm}^k\right) - T\right\} \end{aligned} \tag{22}$$

$$= u_1(y,t) + u_2(\mathbf{s},\mathbf{z}) - \lambda(y + p_i - Y) - \sigma(t - T)$$
$$- \sum_{jk} s_j^k c_{ij}^k - \sum_k z^k c_i^k.$$

In the last expression, we have used (13) and (21) to collect the expected values of the travel costs, travel times and direct utilities into the expected generalized costs ${c^k}_{ij}$ and ${c^k}_i$.

The optimal conditions become

$$\frac{\partial L}{\partial y} = 0 \Rightarrow \lambda = \frac{\partial u_1(y,t)}{\partial y} \tag{23}$$

$$\frac{\partial L}{\partial t} = 0 \Rightarrow \sigma = \frac{\partial u_1(y,t)}{\partial t} \tag{24}$$

$$\frac{\partial L}{\partial z^k} = 0 \Rightarrow \frac{\partial u_2(\mathbf{s},\mathbf{z})}{\partial z^k} = c_i^k \quad \forall k \tag{25}$$

$$\frac{\partial L}{\partial s_j^k} = 0 \Rightarrow \frac{\partial u_2(\mathbf{s},\mathbf{z})}{\partial s_j^k} = c_{ij}^k \quad \forall j,k \tag{26}$$

$$\frac{\partial L}{\partial \lambda} = 0 \Rightarrow y + p_i + \sum_{kmj} \left(z^k P_{mj|i}^k + s_j^k P_{m|ji}^k \right) b_{ijm} = Y \tag{27}$$

$$\frac{\partial L}{\partial \sigma} = 0 \Rightarrow t + \sum_{kmj} \left(z^k P_{mj|i}^k + s_j^k P_{m|ji}^k \right) t_{ijm} = T. \tag{28}$$

Note that the derivatives with respect to the lagrange parameters λ and σ (27)–(28) become variants of the original time and budget constraints (7)–(8) where the expectation values have been replaced by explicit expressions.

Let $\mathbf{s}^*$, $\mathbf{z}^*$ be the optimal activity pattern from (23)–(28), and $\mathbf{x}^*$ the implied expected travel pattern (through (12)). Plugging the solution into the objective function, we obtain u_i, the locational utility of i:

$$\begin{aligned} u_i &= u_3(\mathbf{K}_i) + u_1\left(Y - p_i - \sum_{kmj} b_{ijm} x_{ijm}^{k*}, T - \sum_{kmj} t_{ijm} x_{ijm}^{k*} \right) \\ &\quad + u_2(\mathbf{s}^*, \mathbf{z}^*) + E\left(\sum_{kmj} Q_{ijm}^k X_{ijm}^{k*} \right) \\ &= u_3(\mathbf{K}_i) + u_1\left(Y - p_i - \sum_{kmj} b_{ijm} x_{ijm}^{k*}, T - \sum_{kmj} t_{ijm} x_{ijm}^{k*} \right) \\ &\quad + u_2(\mathbf{s}^*, \mathbf{z}^*) - \sum_k z^{k*} c_i^k - \sum_k s_j^{k*} c_{ij}^k + \sum_{kmj} x_{ijm}^{k*} \left(\lambda b_{ijm} + \sigma t_{ijm} \right). \end{aligned} \tag{29}$$

Now, if total travel costs and travel times are small compared to total available income and time, we can make a Taylor expansion of $u_1(y,t)$ around $(Y\text{-}p_i,T)$, which gives us

$$\begin{aligned} u_i \approx\ & u_3(\mathbf{K}_i) + u_1(Y - p_i, T) + u_2(\mathbf{s}^*,\mathbf{z}^*) - \sum_k z^{k*} c_i^k - \sum_k s_j^{k*} c_{ij}^k \\ & + \sum_{kmj} x_{ijm}^{k*} \left[\left(\lambda - \frac{\partial u_1}{\partial y}(Y - p_i, T) \right) b_{ijm} + \left(\sigma - \frac{\partial u_1}{\partial t}(Y - p_i, T) \right) t_{ijm} \right]. \end{aligned} \tag{30}$$

From the optimality conditions, we know that $\lambda = \frac{\partial u_1(y,t)}{\partial y}$ and $\sigma = \frac{\partial u_1(y,t)}{\partial t}$. If the travel costs and travel times are small compared to total income and time, then (y,t) will be close to $(Y\text{-}p_i,T)$ and consequently $\lambda \approx \frac{\partial u_1(Y-p_i,T)}{\partial y}$ and $\sigma \approx \frac{\partial u_1(Y-p_i,T)}{\partial t}$, if u_1 is not highly non-linear for the range of interest. Assuming this, u_i simplifies to

$$u_i = u_3(\mathbf{K}_i) + u_1(Y - p_i, T) + u_3(s^*, z^*) - \sum_k z^{k*} c_i^k - \sum_k s_j^{k*} c_{ij}^k. \tag{31}$$

This is the form of locational utility that we will use in our application.

Note that the utility derived from travel has a logical structure: for each activity type, the utility consists of the utility of the optimal amount of activities minus the number of activities times their expected generalized cost. For perfectly inelastic activity frequencies, like perhaps work, only the last term will matter, since the first one will be constant. Since the expected generalized costs ($c^k{}_i$ and $c^k{}_{ij}$) are (almost[6]) the usual logsum measures, the common approach of just including a workplace accessibility measure in the form of a logsum can be motivated by the framework presented here. If we believe that location i is chosen conditional on work place j, we should use $c^w{}_{ij}$ (where "w" means "work"); if we believe that work place is instead chosen conditional on location, we should use $c^w{}_i$. Using this approach of course means assuming both that work trip frequency really is inelastic and that we can neglect the influence of other trip types.

As explained above, the generalized costs $c^k{}_i$ and $c^k{}_{ij}$ as well as the optimal trip frequencies $s^k{}_j$ and z^k depend on residual time and income, which is defined in (27) and (28). This is because the time and cost sensitivities in the generalized cost, λ and σ, are defined by (23) and (24). This means that not only will households with less available time T and income Y experience higher generalized costs and lower optimal trip frequencies; the larger the expected amount of time and money spent on travel is, the larger the time and cost sensitivities will be. This means that both having relatively less available time or income and having relatively more expected travel expenses (in terms of time and/or money) will tend to increase the importance of good accessibility.

[6]The difference from the most commonly used "logsum" measures is that the time and cost sensitivities (λ and σ) are not constant.

3 Estimating the Travel Model

The estimation of the travel model is described in detail in Eliasson (2000). Here, we will only briefly reiterate what is necessary to follow the discussion of the interaction between the travel model and the location model.

Ideally, the entire model (travel and location choices) should be estimated jointly. This is perfectly possible in principle, but unfortunately, the data we have available does not permit this. The data we will use is on one hand the recent Swedish national travel survey, and on the other hand a recent housing survey for the Stockholm region. The travel survey contains one-day travel diaries, travel supply data such as travel times and costs between pairs of zones, and finally destination supply data such as number of workplaces of different types in each zone. The housing survey contains detailed data on a number of dwellings, together with detailed data on the household living in it. Unfortunately, there are key variables that are only present in one of the data sets, not both. For example, there are no employment data in the housing survey, neither on employment status or profession, nor on workplace. Further, the geographical resolution is lower (sometimes much lower) than in the travel survey, and there is no travel-related data at all; one could for example have hoped to at least get data on car ownership or license holding. In the travel survey, the most important missing variables are housing price and housing characteristics. Income data is also of a much lower quality than in the housing survey (where it has been taken from national income tax registers). Altogether, this means that we have basically no other choice than to estimate the model in two steps.

Another problem was that we could not use precisely the same household segmentation in the travel model as in the location model, due to missing socio-economic variables in each of the two data sets. The segmentation used in the travel model is found in Table 1. The segmentation used in the location model is presented in the next section.

Average trip frequencies for the seven different trip types and five socio-economic segments is found in Table 1. The estimation is made in a way that ensures that average trip frequencies per segment is replicated by the model.

Up to now, we have not specified the functions in the framework. Introducing household segment index h, we use the following specifications:

$$u_1(y, t) = \alpha \ln y + \beta \ln t \tag{32}$$

Table 1 Trip frequencie*s* per day and individual

	Work	School	Service	Other	Social	Leisure	Shopping	Total
Employed	0.939	0.002	0.054	0.341	0.069	0.105	0.061	1.571
At home	0.024	0.002	0.120	0.400	0.089	0.192	0.073	0.899
Students	0.070	0.431	0.034	0.536	0.039	0.067	0.056	1.232
Retired	0.010	0.001	0.073	0.326	0.040	0.182	0.064	0.695
Other	0.068	0.037	0.133	0.408	0.083	0.108	0.063	0.900

$$u_2^h(\mathbf{s},\mathbf{z}) = \sum_k \gamma^{hk}\rho^k \ln\left(z^k\right) + \sum_{jk} \gamma^{hk}\left(1-\rho^k\right)w_j^k \ln\left(s_j^k\right) \tag{33}$$

$$u_3^h(K_i) = \sum_r \theta^{hr} k_i^r. \tag{34}$$

α, β, the θ^{hr}:s, the ρ^k:s and the γ^{hk}:s are parameters to be estimated. $w^k{}_j$ is a measure of the relative size or attractiveness (both interpretations are possible) of destination j with respect to k-trips, normalized such that $\Sigma_j w^k{}_j = 1$ for all k. $k^r{}_i$ are the elements of the $\mathbf{K}_i$ vector, measuring characteristics such as floorspace, number of rooms and so on. The ρ^k parameters are mixing parameters, measuring the fraction of k-trips that are location-generic.

With these specifications, we can solve the expected travel pattern from (23) to (28):

$$x_{jm}^k = E\left(X_{jm}^k\right) = \gamma^k \frac{\left(1-\rho^k\right)w_j^k}{c_{ij}^k} P_{m|ji}^k + \gamma^k \frac{\rho^k}{c_i^k} P_{mj|i}^k. \tag{35}$$

The strict generalized cost $c^k{}_{ijm}$ from (13) becomes

$$c_{ijm}^k = \frac{\alpha}{y} b_{ijm} + \frac{\beta}{t} t_{ijm} + q_{ijm}^k \tag{36}$$

and the choice probabilities and expected generalized costs are found in (16)–(20). Note that y and t in (35) are calculated endogenously through the budget constraints (27)–(28):

$$y = Y - p_i - \sum_{kmj} \left(z^k P_{mj|i}^k + s_j^k P_{m|ji}^k\right) b_{ijm} \tag{37}$$

$$t = T - t - \sum_{kmj} \left(z^k P_{mj|i}^k + s_j^k P_{m|ji}^k\right) t_{ijm}. \tag{38}$$

The direct trip disutilities $q^k{}_{ijm}$ are assumed to be linear functions of trip characteristics (waiting times, in-vehicle times etc.) and socioeconomic characteristics (license holding, sex etc) and include mode- and trip type-specific constants.

These are admittedly simple suggestions. There are two advantages with deliberately choosing simple functional forms at this stage, however. First, it makes the estimation process simpler, both because the functions are linear in their parameters and because we can obtain analytical solutions to the optimal travel pattern (23)–(28). Since this is a fairly large and complex model anyway, it seems desirable to reduce the computational burden and complexity whenever possible, at least in this first application. Second, showing that the framework works reasonably well even with such simple functional forms is a convincing argument for the framework as such. Certainly, it will work even better with more flexible functional forms, say Box-Cox, CES or translog functions. This is a natural next step once we have established the applicability of the approach.

For work and school trips, we assume that the trip frequencies are constant within a household segment, and equal to γ^w and γ^s, respectively. Workplace and school location is assumed to be chosen conditional on residence location, so this is not determined when choosing residence[7]. Since it can be assumed that there is no real "incentive for variation" over workplaces and schools, work and school trips should be assumed to be purely location-generic, and we set the mixing parameters $\rho^w = \rho^s = 1$. This means that the expected travel pattern for work trips becomes

$$x_{jm}^{work} = \gamma^w P_{mj|i}^{work}, \tag{39}$$

where $P^{work}{}_{jm|i}$ is the choice probability from (16). The school trip pattern is obtained in the same way.

4 Estimating the Location Choice Model

4.1 The Location Choice Probability

The location utility u_i was derived above in (31). We assume that we can only observe u_i up to an additive constant ε_i, which is different even across seemingly similar households. Using the function specifications above (32)–(34), we obtain the following expression for a household from segment h:

$$\begin{aligned} u_i = & \sum_r \theta^{hr} k_i^r + \alpha \ln(Y - p_i) \\ & - \sum_k \gamma^{hk} \rho^k \ln c_i^k - \sum_{jk} \gamma^{hk} \left(1 - \rho^k\right) w_j^k \ln c_{ij}^k - \gamma^{hw} c_i^w - \gamma^{hs} c_i^s + \varepsilon_i. \end{aligned} \tag{40}$$

$c^w{}_i$ and $c^s{}_i$ are the expected generalized cost for work and school trips. $c^k{}_i$ and $c^k{}_{ij}$ are expected generalized costs for location-generic and location-specific discretionary trip types, respectively.

This particularly simple form comes from choosing the travel utility function $u_2(\mathbf{s},\mathbf{z})$ to be logarithmic, which causes the terms $\Sigma_k z^k c^k{}_i$ and $\Sigma_{jk} s^k{}_j c^k{}_{ij}$ in (31) to be independent of i, and thus cancel out when comparing different residences i.

From the travel model estimation, we obtain the ρ^k:s and the sizes of the γ^{hk} relative to each other, i.e. we know each γ^{hk} up to a multiplicative constant. That we

[7]It would be interesting to compare this assumption with the opposite one, that location of residence is chosen conditional on work place. Unfortunately, our data does not permit this, since the housing survey does not contain data on workplaces.

only know γ^{hk} up to a multiplicative constant is because we cannot separate the variance of ε_i from the γ^{hk} parameters. It remains to estimate α, γ^w, γ^s, the θ^{hr}:s and the scale of the γ^{hk}:s.

Assuming that the idiosyncratic term ε_i is generalized extreme value distributed, we obtain the following expression for the probability that a household will choose residence i. Dividing the residences in a number of disjoint sets H_a, we write

$$P_i = \frac{\exp(\mu_4 U_a)}{\sum_a \exp(\mu_4 U_a)} \frac{\exp(u_i)}{\sum_{i \in H_a} \exp(u_i)}, \tag{41}$$

where $U_a = \ln \sum_{i \in H_a} \exp(u_i)$.

As subsets H_a, we will use rented apartments, owned apartments and houses (which are always owned). Other divisions are of course possible. In Sweden, the price on the former apartment type is subject to rent control, while the price of the latter is set on the open market, operating much like a standard house market. That the latter apartment type is “owned” is not the whole truth. The legal status of “owned” apartments in Sweden is a bit complicated; for our purposes, it is sufficient to know that although the apartment is “owned” in the sense that the contract of the apartment is bought on an open market, a “rent” is also paid to the “association” owning the building, covering much of the maintenance costs and often also heating and water.

4.2 Data and Estimation Methodology

The data on dwellings and residents came from the Swedish Housing Survey, constituting a representative sample (after weighting) of households and dwellings in the county of Stockholm. It is thus not a sample of only recent movers, which might introduce a certain inconsistency in the material since we use the *current* prices and accessibilities in the estimation – not their values when the location choice was taken. We have, however, no data on how long the household has been living in their dwelling.

There were 919 houses, 1,242 owned apartments and 882 rented apartments in the data set. Clearly, this is too many alternatives to handle, so we used choice set sampling to reduce the number of terms in the second denominator of (41). McFadden (1978) proved that a multinomial logit model with sampled choice sets gives asymptotically unbiased coefficient estimates, if each sampled alternative is weighted with its sampling probability, or if alternative-specific constants are used. We sampled 20 residence alternatives for each observation, chosen to obtain seven houses, seven owned apartments and seven rented apartments in each choice set (including the chosen residence). Having the different housing types in different nests has the additional benefit of eliminating the sampling weights, since they cancel out in each nest.

The population are divided into seven different socioeconomic segments. These are:

Segment number	Description
1	Retired, single
2	Working, single, no children
3	Working, single, one or more children
4	Retired, married
5	Working, married, no children
6	Working, married, one child
7	Working, married, two or more children

"Married" is just a shorthand for two-adult households; whether these are actually married or just living together does not matter. "Working" means that at least one adult in the household is working, unemployed or studying. "Retired" means that all adults in the households are retired. There were no retired households with children in the sample.

After testing several different specifications and variables, we arrived at the following model derived from (40).

$$u_i = \theta^h{}_a + \theta_{fs} * [floorspace] + \theta^h{}_{room} * [roomdensity] + \alpha \ln(Y - p_i) - \Gamma * [discr.acc] - \gamma^{hw} c^w{}_i - \gamma^{hs} c^s{}_i. \tag{42}$$

$\theta^h{}_a$ are constants to be estimated, different for the three dwelling types, rented apartments, owned apartments and houses, and different for each household segment. The constant for houses was normalized to 0.

"floorspace" is the floorspace in square meters, and θ_{fs} is a parameter to be estimated. We tested letting this be segment-specific, but these differences were not significant. Y, income, is the available income per month after tax and various supports, and p_i is the average monthly cost for dwelling i. For rented apartments, this is simply rent per month. For owned apartments, it is the sum of mortgages, rent and maintenance costs (to the extent these are not covered by rent). For houses, it is the sum of mortgages and maintenance costs (including heating, water etc.). All costs were reported by the households themselves and included in the survey data.

"Room density" deserves an explanation. It turned out that the number of rooms was so correlated with floorspace that we could not estimate both parameters for floorspace and number of rooms. Instead, we constructed a measure of whether a dwelling had few or many rooms relative to its floorspace, in an approach somewhat similar to two-stage least squares. To our knowledge, this way of separating floorspace and number of rooms has not been proposed before.

First, we estimated a linear regression model, regressing the number of rooms (*#rooms*) on floorspace:

$$\#rooms_i = \alpha_0 + \alpha_1 * floorspace_i. \tag{43}$$

Then we defined "room density" to be the residuals from this model, i.e.

$$[room\ density]_i = \#rooms_i - (\ \alpha_0 + \ \alpha_1 * floorspace_i). \tag{44}$$

This "relative" room measure was then entered as an explanatory variable, and a separate parameter for each segment was estimated – the parameter for floorspace was not significantly different between segments. Two other methods we tried were using *floorspace/#rooms* and *#rooms/floorspace*, but neither of them worked well.

"*discr. acc*." is the accessibility of discretionary trips, i.e. all trip types except work and school trips. As can be seen from (39), this is defined to be

$$discr.acc = -\sum_k \gamma^{hk} \rho^k \ln c_i^k - \sum_{jk} \gamma^{hk} \left(1 - \rho^k\right) w_j^k \ln c_{ij}^k. \tag{45}$$

All the parameters in this expression are already obtained from the travel model. Note from the definition of the generalized costs $c^k{}_i$ and $c^k{}_{ij}$ that these depend on residual income (available income minus housing and expected total travel expenses) and residual time (available time minus expected total travel time). These will therefore be different across households. The parameter Γ is the same for all household segments. We tested relaxing this, but this turned out to yield insignificant parameter values, sometimes with the wrong sign. This was actually expected, since differences in accessibility preference across household segments should (theoretically) be captured by the γ^{hk} parameters. We also tried to estimate all the parameters $\Gamma\gamma^{hk}$ without constraints, to see if we obtained the same ratios between γ^{hk} as we did from the travel model. Unfortunately, the $c^k{}_i$:s and $c^k{}_{ij}$:s were so correlated across k that this was not possible.

The working trip parameters γ^{hw} are assumed to be equal for all household segments with workers, and zero for others.

4.3 Estimation Results

Results from the estimation are found in the Table 2.

The basic preferences for housing type – rented apartment, owned apartment or house – per household segment are what we expected (remember that the "house" constant is normalized to 0). All else being equal, single-adult households (segment 1–3) tend to prefer (or at least live in) rented apartments, followed by owned apartments and houses last. The "preference" (if we interpret the constants as preference parameters) for rented apartments over owned ones is especially strong for segment 3, retired singles. Two-adult households tend to prefer houses, followed by rented apartments with owned apartments last. Especially couples with two or more children tend to avoid owned apartments. This indicates that households from this segment prefer houses if they can afford to buy a dwelling at all; if they cannot afford to buy, renting an apartment is the only choice.

Table 2 Parameters in the locational utility, Log-likelihood 8,398.51, ρ^2 0.0671

Variable	Estimate	Std. Error	t-ratio
θ^h_a Rented apt. segm. 1	2.636	0.54	4.9
θ^h_a Owned apt. segm. 1	2.044	0.46	4.4
θ^h_a Rented apt. segm. 2	3.009	0.54	5.6
θ^h_a Owned apt. segm. 2	2.667	0.49	5.5
θ^h_a Rented apt. segm. 3	3.096	0.66	4.7
θ^h_a Owned apt. segm. 3	1.372	0.54	2.6
θ^h_a Rented apt. segm. 4	−0.076	0.18	−0.4
θ^h_a Owned apt. segm. 4	−0.175	0.19	−0.9
θ^h_a Rented apt. segm. 5	−0.438	0.14	−3.1
θ^h_a Owned apt. segm. 5	−0.820	0.19	−4.4
θ^h_a Rented apt. segm. 6	−0.486	0.19	−2.6
θ^h_a Owned apt. segm. 6	−1.008	0.25	−4.0
θ^h_a Rented apt. segm. 7	−0.343	0.15	−2.3
θ^h_a Owned apt. segm. 7	−1.161	0.25	−4.6
Work trips (γ^w)	0.151	0.05	3.3
Discr. trips (Γ)	1.029	0.29	3.5
Income (α)	0.587	0.08	7.1
Floorspace (θ_{fs})	0.00418	0.00058	7.2
Room density segm. 1	−0.357	0.09	−3.8
Room density segm. 2	−0.689	0.07	−9.6
Room density segm. 3	0.408	0.14	3.0
Room density segm. 4	0.258	0.08	3.2
Room density segm. 5	0.135	0.05	2.6
Room density segm. 6	0.273	0.08	3.6
Room density segm. 7	0.491	0.06	7.9
Logsum (μ_4)	0.7579	0.120	6.3

That the work trip parameter γ^w is positive and significant shows that accessibility to jobs indeed has a significant influence on the attractiveness of a location, which is what we expected. Further, this also holds for Γ, the parameter measuring the utility derived from discretionary trips. Including accessibility to other things than workplaces thus contributes significantly to the model's explaining power. However, experimenting shows that including *several* discretionary trip types in the location utility function contributes relatively little beyond the first few; once we have included, say, shopping trips, adding for example service trips increases the log-likelihood value only marginally. The reason for this is of course that the generalized costs for service and shopping trips (in this example) are highly correlated, which is natural since service and shopping establishments tend to locate near each other. In an urban region where different types of establishments do not tend to locate close to each other to the same extent as in Stockholm, the generalized costs will consequently be less correlated – one could for example imagine a city where shops and leisure establishments (like restaurants and cinemas) were more geographically separated than in Stockholm.

Zones do not differ much with respect to accessibility to schools: most school trips are short (within the zone), so school accessibility largely depends merely on the number of schools within the zone – which in turn is highly correlated with workplace and population density, and thus correlates with workplace accessibility. Hence, the school trip parameter γ^{hs} did not come out significantly different from zero, so it was excluded from the model.

That the income parameter α is positive is of course necessary if we want to put some trust in our model. We just point out that obtaining this result is not entirely trivial, since the price may be well predicted by the other included variables. We recall the discussion in the introduction (Sect. 1.4) on the determination of market prices; were all households identical, we would expect prices to be determined such that utility was the same everywhere, implying that the price would be perfectly predicted by the other variables.

Turning to the preferences for rooms and floorspace, we see the expected results. All segments value more floorspace higher than less, all else being equal. The differences in floorspace preference between household segments were not significant, so we estimated a joint parameter for all segments. The room density, as explained above, measures the relative preference for many rooms conditional on the amount of floorspace. A negative parameter thus reveals that the segment prefers fewer rooms than the average number of rooms implied by the floorspace, and conversely a positive parameter reveals a preference for more rooms. If we order the segments with the segment wanting the most rooms first, we get

Segment no.	Description	Room preference
7	Working, married, $\geq$2 children	0.491
3	Working, single, $\geq$1 child	0.408
6	Working, married, one child	0.273
4	Retired, married	0.258
5	Working, married, no children	0.135
1	Retired, single	−0.357
2	Working, single, no children	−0.689

The results seem intuitive, indicating that this way of separating floorspace and number of rooms in fact captures the room preference the way it should. That retired have higher room preference than working households probably depends on that, in many cases, their current dwelling was bought when the now retired household had children living at home.

5 Conclusions

The focus of this paper is on modeling the influence of accessibility on the household's location decision. Our main theoretical contribution is an elaborate specification of what we should mean by "accessibility" in this context. This is done by letting the utility of a location consist of four parts: the indirect utility of income and time net of housing costs and expected total travel times and travel costs, the direct utility of the optimal activity pattern, the direct disutility of the expected travel pattern and the direct utility derived from location characteristics. We have also shown how a non-linear indirect utility function can be introduced in a consistent way into the joint activity pattern-location choice problem. Letting the indirect utility function be non-linear is a prerequisite for the available income and time to influence the choice of location, activities and travel.

In the empirical part of the paper, we have presented methodology and results from the estimation of an instance of the framework – TILT, Tool for Integrated analysis of Location and Travel, a land use-transportation model for the Stockholm region. TILT is an operational housing demand model, although the presentation here mainly serves to show that the framework is possible to estimate and yields the expected results. We found that the attractiveness of a location increased both with the accessibility to workplaces and with the accessibility to service, shops etc. Although this was expected, we want to point out that it is fairly common in location choice models to include only workplace accessibility, a practice that our results indicate will give biased estimations of location utilities. Further, we demonstrated the construction of a measure of the relative number of rooms given the floorspace. Including this measure in the locational utility yielded satisfactory results.

Acknowledgments Thanks to Folke Snickars and Lars-Göran Mattsson for valuable comments and to Staffan Algers for help with the data. Financial support from the Swedish Transport and Communications Research Board is gratefully acknowledged.

References

Abraham JE, Hunt JD (1997) Specification and estimation of a nested logit model of home, workplaces and commuter mode choices by multiple worker households. Transp Res Rec 1606:17–24

Algers S, Eliasson J, Mattsson L-G (2005) Is it time to use activity-based urban transport models? A discussion of planning needs and modelling possibilities. Ann Reg Sci 39:767–789

Alonso W (1964) Location and land use. Harvard University Press, Cambridge

Anas A (1982) Residential location markets and urban transportation. Academic Press, New York

Anas A (1995) Capitalization of urban travel improvements into residential and commercial real estate: simulations of a unified model of housing, travel mode and shopping choices. J Reg Sci 35:351–375

Anas A, Moses L (1979) Mode choice, transport structure and urban land use. J Urban Econ 6:228–246

Axhausen K, Gärling T (1992) Activity-based approaches to travel analysis: conceptual frameworks, models, and research problems. Transp Rev 12:323–341

Becker G (1965) A theory of the allocation of time. Econ J 75:493–517

Ben-Akiva M, Bowman JL (1998) Integration of an activity-based model system and a residential location model. Urban Stud 35:1131–1153

Ben-Akiva M, Lerman S (1985) Discrete choice modelling. MIT, Cambridge

Chang JS (2006) Models of the relationship between transport and land-use: a review. Transp Rev 26(3):325–350

De Serpa A (1971) A theory of the economics of time. Econ J 81:828–846

Eliasson J (2000) Integrated travel pattern modeling. Theory and estimation of an integrated land use-transportation model. In Eliasson J (ed) Transport and location analysis, Dissertation, Department of Infrastructure and Planning, Royal Institute of Technology, Stockholm

Eliasson J, Mattsson L-G (1998) Modeling long-period trip patterns with time and money constraints – the TILT land-use/transportation model. In: Meersman H, van de Voorde E and Winkelmans W (eds.): World transport research – selected proceedings from the 8th World Conference on Transport Tesearch, Antwerp, Belgium, 12–17 July 1998. Amsterdam, Pergamon

Eliasson J, Mattsson L-G (2000) A model for integrated analysis of household location and travel choices. Transp Res A 34:375–394

Ellickson B (1981) An alternative test of the hedonic theory of housing markets. J Urban Econ 9:56–79

Ettema D, Timmermans H (eds) (1997) Activity-based approaches to travel analysis. Pergamon, Oxford

Evans A (1972) On the theory of the valuation and allocation of time. Scott J Polit Econ 19:1–17

Evans A (1973) The economics of residential location. MacMillan, London

Fujita M (1989) Urban economic theory, land use and city size. Cambridge University Press, Cambridge

Hunt JD, Kriger DS, Miller EJ (2005) Current operational urban land-use-transport modelling frameworks: a review. Transp Rev 25(3):329–376

Jara-Díaz SR, Martínez FJ (1999) On the specification of indirect utility and willingness to pay for discrete location models. J Reg Sci 39:675–688

Jonsson D (2008) Analysing sustainable urban transport and land-use. PhD dissertation, Royal Institute of Technology, Stockholm

Martínez F (1992) The bid-choice land use model: an integrated economic framework. Environ Plann A 24:871–885

Martínez F (1996) MUSSA: A land use model for Santiago city. Transp Res Rec 1552:126–134

McFadden D (1974) Conditional logit analysis of qualitative choice behaviour. In: Zarembka P (ed) Frontiers in econometrics. Academic Press, New York

McFadden D (1978) Modeling the choice of residential location. In: Karlqvist A, Lundqvist L, Snickars F, Weibull J (eds) Spatial interaction theory and planning models. North Holland, Amsterdam

Mills ES (1967) An aggregative model of resource allocation in a metropolitan area. Am Econ Rev 57:197–210

Muth RF (1969) Cities and housing. The University of Chicago Press, Chicago

Rosen S (1974) Hedonic prices and implicit markets: product differentiation in pure competetition. J Polit Econ 82:34–55

Swedish Consumer Agency [Konsumentverket] (1997) Koll på pengarna? Available at www.konsumentverket.se

Waddell P (2000) A behavioral simulation model for metropolitan policy analysis and planning: residential location and housing market components of UrbanSim. Environ Plann B 27:247–263

Waddell P (2002) UrbanSim: modeling urban development for land use, transportation and environmental planning. J Am Plann Assoc 68(3):297–314

Wegener M (1994a) Operational urban models – state of the art. J Am Plann Assoc 60:17–29

Wegener M (1994b) Models of urban land use, transport and environment. In: Advances in Spatial Science. Springer, Berlin

Wegener, M. (1998) Applied models of urban land use, transport and environment: state of the art and future developments. In: Lundqvist L, Mattsson L-G, Kim TJ (Eds.), Network Infrastructure and the Urban Environment. Springer, Berlin

Wegener M (2004) Overview of land-use transport models. In: Hensher D et al (eds) Handbook in transport, vol 5. Pergamon, Kidlington

Wheaton WC (1977) A bid rent approach to housing demand. J Urban Econ 4:200–217

Modeling Residential Location in UrbanSim

Paul Waddell

Abstract This chapter provides a description of the residential location component of UrbanSim, drawing on applications of UrbanSim in numerous metropolitan areas. The first section provides an overview of the UrbanSim system, with particular attention to the role of the residential location choice model within it. The second section describes the Open Platform for Urban Simulation, and explains how choice models in general, and more specifically residential location choice models, are created in this framework. The third section provides a comparison of recent applications of the UrbanSim residential location choice framework, along with lessons learned. The final section summarizes the current status of the model system and outlines current development efforts.

1 Introduction

This chapter provides a description of the residential location component of UrbanSim, drawing on applications of UrbanSim in metropolitan areas as diverse as Amsterdam, Detroit, Paris, Phoenix, Salt Lake City, San Francisco, Seattle and Zürich to highlight the evolution of the framework through practical application in a variety of settings. The first section provides an overview of the UrbanSim system, with particular attention to the role of the residential location choice model within it. The second section describes the Open Platform for Urban Simulation, and explains how choice models in general, and more specifically residential location choice models, are created in this framework. The third section provides a comparison of recent applications of the UrbanSim residential location choice framework, along with lessons learned. The final section summarizes the current status of the model system and outlines current development efforts.

P. Waddell
Department of City and Regional Planning, University of California, Berkeley. Berkeley, CA, USA
e-mail: waddell@berkeley.edu

F. Pagliara et al. (eds.), *Residential Location Choice*, Advances in Spatial Science,
DOI 10.1007/978-3-642-12788-5_8,

2 Overview of UrbanSim

UrbanSim was designed initially in the late 1990s to respond to a perceived gap in operational models to support metropolitan-scale coordination of transportation and land use planning and analysis (Waddell 2000, 2002; Waddell et al. 2003). Metropolitan Planning Agencies needed models to assess the consequences of alternative transportation plans and policies on urban development and travel patterns. Some wanted to evaluate the effects of land policies such as the use of urban growth boundaries, or policies to promote transit-oriented development. Most wanted to be able to address these kinds of policy analysis questions with models that were behaviorally clear and as transparent as possible, avoiding the problems identified three decades ago by Lee's critical assessment of the state of large scale urban simulation (Lee 1973), and the more general skepticism of "black-box" models that were so complex that their logic could not be explained to policy-makers or the public.

The original design of UrbanSim adopted several elements to address these modeling requirements, and these have remained foundational in the development of the system over time. These design elements include:

- The representation of individual agents: initially households and firms, and later, persons and jobs.
- The representation of the supply and characteristics of land and of real estate development, at a fine spatial scale: initially a mixture of parcels and zones, later gridcells of user-specified resolution.
- The adoption of a dynamic perspective of time, with the simulation proceeding in annual steps, and the urban system evolving in a path dependent manner.
- The use of real estate markets as a central organizing focus, with consumer choices and supplier choices explicitly represented, as well as the resulting effects on real estate prices. The relationship of agents to real estate tied to specific locations provided a clean accounting of space and its use.
- The use of standard discrete choice models to represent the choices made by households and firms and developers (principally location choices). This has relied principally on the traditional Multinomial Logit (MNL) specification, to date.
- Integration of the urban simulation system with existing transportation model systems, to obtain information used to compute accessibilities and their influence on location choices, and to provide the raw inputs to the travel models.
- The adoption of an Open Source licensing for the software, written originally in Java, and recently reimplemented using the Python language. The system has been updated and released continually on the web since 1998 at www.urbansim.org.

Following the original design of UrbanSim, and the implementation of a working prototype of the model in the Eugene-Springfield metropolitan area (Waddell 1998, 2000, 2002), many of these elements have been adopted by other model systems, including the Oregon2 model framework and Delta.

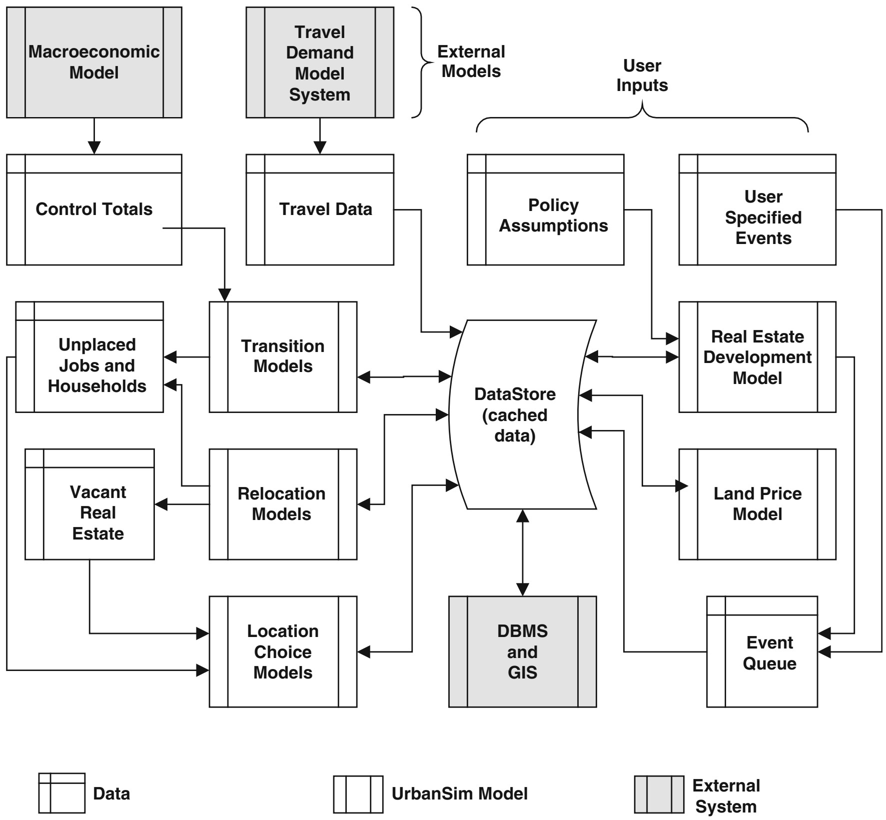


Fig. 1 UrbanSim model system

Figure 1 summarizes the overall flow of the model system in a typical application, including its inputs and interactions. Note that there are two-way interfaces to the travel model system, and a one-way interface with external macroeconomic models that predict the overall pattern of economic growth. In addition, users specify assumptions such as how comprehensive land use plans will be used to constrain the patterns of real estate development.

3 The Household Location Choice Model

The UrbanSim model system contains model components representing household and employment relocation and location choices, and real estate development and prices has been described in previous papers using a range of specifications and locations (Waddell 2000, 2002). This chapter, in keeping with the focus of this volume, focuses on only one component of the model system: household location.

The function of the household location choice model is straightforward, as is the data structure on which it operates. A list of households, generated using a synthetic population synthesizer (Beckman et al. 1996), is represented in the base year database as a table with one row per household. Each household has a unique identifier, attributes such as number of persons, income, number of workers, presence of children, and a unique identifier for its location. As the model system proceeds in the first simulation year, the demographic transition model adds new households to the household table, providing their characteristics and a unique identifier, but not a location identifier. Then the household relocation model simulates the choices of certain households to move from their current location, and resets the location identifiers of the moving households to a null value. As a result of these two models, then, the household table contains some households that have moved into the region, and some that have been predicted to move within the region. These locating households are selected by identifying all households in the table that have a null location identifier. This is the set of households that the location choice model is applied to.

The framework for the household location choice model, like most of the models in the UrbanSim model system, is a standard choice model. Although more sophisticated choice model structures can be used, the most common in practice is the Multinomial Logit Model (MNL) (see McFadden 1973, Ben-Akiva and Lerman 1985, or Train 2003 for a thorough description of the model structure and how it compares to alternatives such as nested or mixed logit).

The underlying logic of the model is that households that are in the market for a location take into consideration their own characteristics, such as income, and household size, and consider a sample of available, vacant housing units and their price and characteristics such as density, age, and accessibility to employment and other opportunities. The relative attractiveness of these alternatives is measured by their utility. The choice model then proceeds to compute the probabilities of making a location choice from the available alternatives, defined as vacant housing units, given the preferences and budget constraints of locating households. Once location probabilities are computed, the predicted choices are simulated, using one of the available algorithms to reflect different assumptions regarding how the housing market clears.

The model proceeds in steps as outlined in Fig. 2. After loading the model specification and coefficients from input data, the model selects the agents that will be making a choice. As noted earlier, this is done in UrbanSim by selecting all households who do not have a location identifier, that is, all households that need a location. Note that the model can be stratified into submodels, reflecting groups of households defined by some characteristic for which the user wishes to estimate the model separately. This can be done by income, or household size, number of workers, or any other household characteristic that the user might use to examine differences in location behavior. In market research this is often referred to as market segmentation.

The next step is the determination of the choice set. The universal choice set defined for this model is the total vacant housing stock. In most metropolitan areas this can be a very large number of housing units, and would be both behaviorally

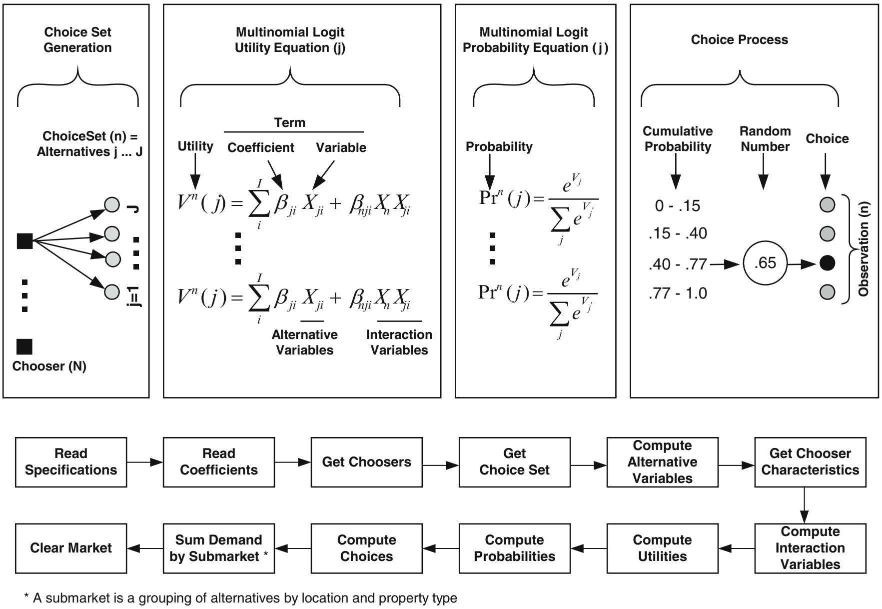


Fig. 2 Computational process for household location choice model

unrealistic to consider, and computationally excessive. A typical household does not exhaustively examine every unit on the market, as the search costs for doing this would be prohibitive in time and effort. UrbanSim does not impose the assumption that they do, and allows users to specify alternative sampling frameworks for the alternatives to be considered. Options currently include random, weighted, and stratified sampling. Corrections for the sampling protocol are needed to ensure that the coefficients of the model are not biased, and UrbanSim contains functionality to make these corrections.

A major consideration in defining the choice set is the question of spatial scale. Housing units are ultimately the elemental basis for the residential location choice. UrbanSim supports modeling residential location choice at the parcel (or building) level, or at more aggregate units of geography such as gridcells, or traffic analysis zones, or neighborhoods or other spatial units. The definition of the geographic unit of analysis is a configuration choice that the user makes in setting up the model system, and is not hard-coded into the software. Different applications of UrbanSim have used large districts, zones, gridcells, and buildings.

Once the choosers and the choice set are determined, the model proceeds to compute the utility for each of the sampled alternatives, for each locating household. Utility calculations involve computing some variables that describe characteristics of the alternatives, and other variables describing interactions between household characteristics and characteristics of alternatives. Alternative characteristics might include such variables as the residential unit density in the neighborhood, or the housing type of the unit, or the access of the location to employment or shopping opportunities. Interaction variables include measures such as the income of the household minus the annualized cost (rent) of the housing unit. One point that bears noting is that household characteristics can only enter a choice model through interaction terms, since otherwise there would be no variation among alternatives and there is no way to estimate a coefficient for such a variable.

Once the variables are all computed, the utility is computed as a simple summation of the products of the variable and coefficient vectors. Given the utility values for each household and each of their sampled alternatives, and assuming a particular (Gumbel) distribution for the error term, it is then computationally simple to predict probabilities using a standard multinomial logit model:

$$P(i) = \frac{e^{V_i}}{\sum_{j}^{J} e^{V_j}}.$$

Once choice probabilities are computed, the model simulates the choices made by agents. There are several algorithms that have been implemented in UrbanSim to reflect alternative assumptions of how the market clearing process works. The most traditional economic assumption to make is that prices will simply adjust to clear the market. That is, if more households are predicted to choose houses in scarce locations then the model would raise the price of those houses, force households to choose again, and repeat this process until the prices clear the market in the sense that each vacant housing unit has no more than one household who would choose that unit. While this is a convenient assumption to make, since it simplifies the model considerably, it is not necessarily very realistic in short-term housing markets where disequilibrium may not be uncommon, and transactions costs are high.

There are many frictions in the housing market that make it less than perfectly liquid, and transactions costs such as the time and effort involved in searching, as well as the fees to the realtor and financial agents involved. As a result, UrbanSim has implemented alternative market algorithms. One of these is a capacity-constrained algorithm that clears the market using a first-come, first-served approach. When a house is selected by a household, a contract is signed and the house is taken off the market, making it unavailable for other households, even if the latter might have bid more. In reality, market clearing is likely to fall somewhere between the pure price adjustment and the lottery market clearing protocols.

A new algorithm implemented in UrbanSim is the constrained choice algorithm developed by de Palma et al. (2007). This approach recognizes that constraints on alternatives do exist in real markets, and that imposing an assumption that prices do all of the work in clearing the market may actually impose a significant bias on the coefficients of the choice model. Using a revised estimation technique that incorporates the effects of availability constraints, this algorithm reduces the bias by attempting to un-mix the price and constraint effects simultaneously. This approach has been tested in the Paris housing market, and shows significant differences from the more conventional assumptions regarding market clearing that ignore the role of constraints.

4 Data Structure and Preparation

The data used in UrbanSim can be as detailed as needed or supported by available data resources, and can therefore vary considerably from one application to another. The most detailed applications of UrbanSim to date are in San Francisco and in Seattle, both of which have been recently applied at the parcel and building level. That is, the unit of analysis for the Household Location Choice Model is the individual residential building, and the associated parcel on which the building is located. Information from travel model zones, including accessibility measures, can be assigned to parcels within a zone, and smaller-scale proximity measures can also be used. An example of the entity–relationship relating households to buildings, and buildings to parcels and other spatial units and entities is shown in Fig. 3, patterned after the San Francisco UrbanSim application, which is integrated with one of the only operational activity-based travel models. The geographies used in the San Francisco application are shown in Fig. 4, with Traffic Analysis Zones for the city, and parcels in a small section.

In the case of such detailed models, it is often a question whether the level of detail is so high that the errors in the data make the model difficult to estimate reasonable parameters for. Similar concerns are often raised about the computational cost of running models at this level of detail. We have found that the model estimation can actually be significantly improved over more aggregate specifications, in spite of errors in the data. We think this is reflective of the closer match between the model specification and the actual entities and behavior in the real world.

The model can be configured to run at more aggregate levels of geography. The most commonly used spatial unit of analysis with UrbanSim is the gridcell, typically using a resolution of 150 m (see, for example, Waddell 2000, 2002; Waddell et al. 2007a), though some applications have used smaller or larger cell sizes. The gridcell geography is quite convenient for certain kinds of operations, such as computing variables that are based on queries of surrounding cells, or for exporting data for visualization in a GIS environment. In other respects, however, gridcells pose problems. For example, the true underlying unit of analysis is the parcel, but gridcells intersect parcels and therefore cannot represent very directly the real estate contents of the cell. This process requires rasterization of parcel data, or GIS overlay operations to combine parcel and gridcell layers using a union operation, and fractionating the parcels before re-aggregating the parcel fragments into gridcell summaries. Further, the representation of land policies becomes problematic, due to the loss of a direct connection to the parcel geography. Nevertheless, this is still a very popular and widely used approach to specifying locations, and has sufficient merits to be worth consideration.

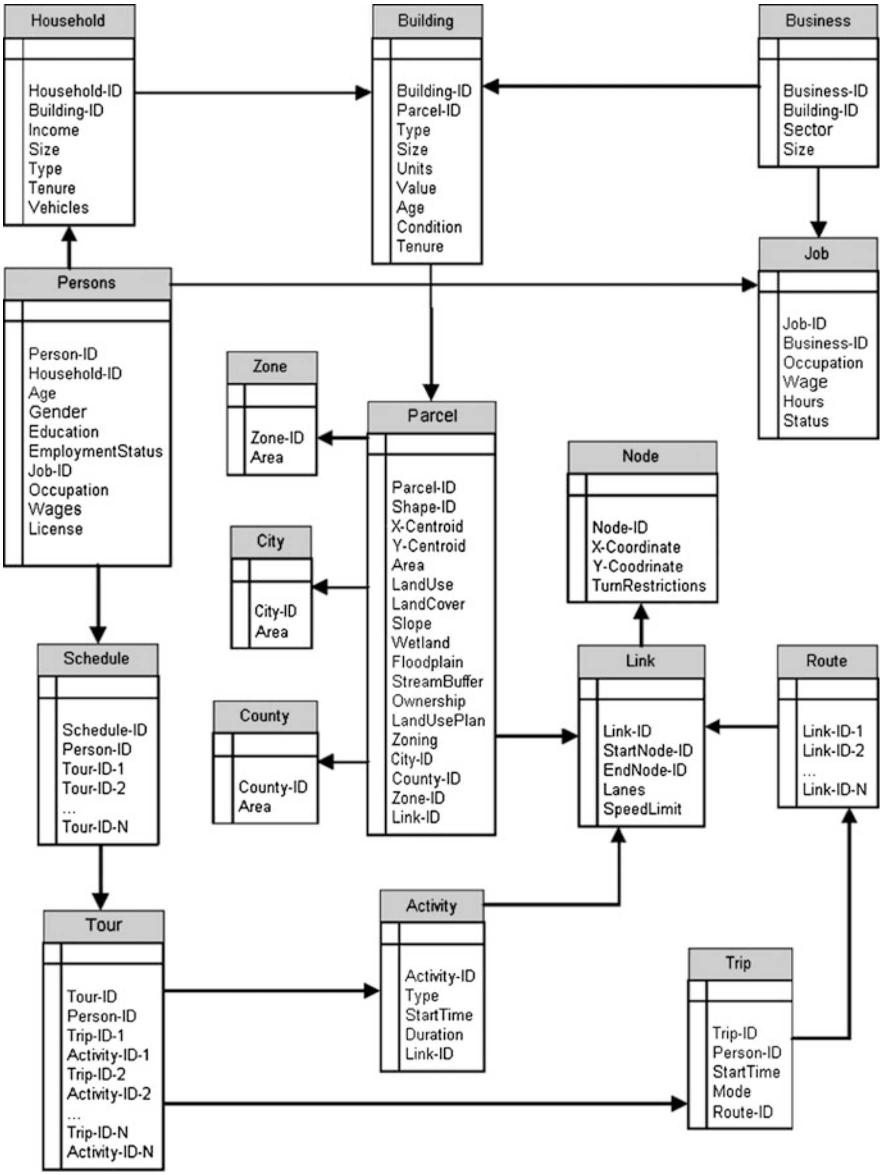


Fig. 3 Entity–relationship diagram for an integrated parcel – activity based travel model system

A third level of application that has been tested is the use of aggregate areas or zones. These could be larger communities such as the 1,300 Communes used in the application of UrbanSim in Paris (de Palma et al. 2007), or smaller districts such as Traffic Analysis Zones, or user-defined neighbourhoods. There are merits and limitations to each of these spatial units of analysis. One thing to note is that it is

Fig. 4 Parcels and traffic analysis zones in San Francisco

easy to use variables representing higher levels of geography in more detailed spatial models, so a multi-level representation of variables influencing location choice is straightforward and reasonably common.

The data preparation for the model system usually begins with acquiring the following data:

- Parcel information from county assessors offices
- Business establishment data from state unemployment insurance records or from private sources such as InfoUSA
- Census data, including both sample microdata and census tract summary tabulations
- Traffic zone geography and travel model results (travel time skims by mode or logsums, by time of day and purpose)
- Environmental features
- Travel survey data

In the development of the San Francisco model, these data were readily available and required modest spatial processing to prepare for use in the model system. Households were available from a synthetic population generator that combines microdata samples with tabulations to produce a synthetic baseline population that is consistent with the census data. The database development for the San Francisco model application required approximately 6 person-months of effort over an 18-month period. In other applications, the data could require considerable additional effort, which might suggest adopting a coarser level of analysis.

5 Specification and Estimation

The specification of the Household Location Choice Model in UrbanSim, like the other choice models in the system, involves creating a specification that includes the chooser and alternative characteristics to be considered in the model. It also involves determining whether to stratify the estimation by some characteristic of the households making location choices. In the San Francisco application, the model was stratified by the number of workers in the household, reflecting the hypothesis that there may be significant differences in their locational preferences.

The variables used in the specification for the San Francisco model, and preliminary results of model estimation are shown in Tables 1 and 2. All the estimation of choice models in UrbanSim is done using Maximum Likelihood Estimation, with integrated estimation software developed as part of the system. The estimation time for the San Francisco models requires less than 30 seconds, and can be iteratively re-specified and re-estimated in seconds during the process of developing a desired model specification.

Accessibility is measured using bus and auto modes, by computing the employment opportunities available within 30 min travel time by each mode in the a.m. peak. Housing price is estimated by imputing an annual rent from the total per-unit

Table 1 Variables used in the San Francisco household location choice model

Variable	Description
ln_emp_30_bus	Natural log of total employment within 30 min by bus mode in the am peak
ln_emp_30_hwy	Natural log of total employment within 30 min by drive alone mode in the am peak
ln_households_in_zone	Natural log of total households in traffic analysis zone
ln_inc_avg_inc	Natural log of household income × the zone average income
ln_inc_building_sf_per_unit	Natural log of household income × average square feet per housing unit on the parcel
ln_inc_minus_cost	Natural log of (household income – annual imputed rent)
ln_inc_sector_3_employment_in_zone	Natural log of household income × zonal retail employment
ln_residential_units	Natural log of residential units, as a size variable

Table 2 Preliminary estimation results from San Francisco

Workers	Variable	Coeff	t-stat
No-workers	ln_emp_30_bus	0.203	12.66
	ln_households_in_zone	0.291	9.60
	ln_inc_avg_inc	0.115	2.58
	ln_inc_building_sf_per_unit	−0.005	−0.75
	ln_inc_minus_cost	0.013	2.26
	ln_inc_sector_3_employment_in_zone	−0.007	−0.73
	ln_residential_units	1.015	73.37
Observations	4,008		
Rho-squared	0.28		
One Worker	ln_emp_30_bus	0.017	0.95
	ln_households_in_zone	0.111	2.86
	ln_inc_avg_inc	0.425	7.19
	ln_inc_building_sf_per_unit	0.001	0.15
	ln_inc_minus_cost	0.040	4.78
	ln_inc_sector_3_employment_in_zone	−0.029	−2.92
	ln_residential_units	0.952	56.90
Observations	2,947		
Rho-squared	0.17		
Two or more workers	ln_emp_30_hwy	0.074	1.77
	ln_households_in_zone	0.015	0.39
	ln_inc_avg_inc	−0.013	−0.25
	ln_inc_building_sf_per_unit	0.016	2.10
	ln_inc_minus_cost	0.064	6.78
	ln_inc_sector_3_employment_in_zone	−0.055	−5.77
	ln_residential_units	0.816	49.85
Observations	3,045		
Rho-squared	0.12		

assessed value of each residential building. The imputed rent is interacted with income to reflect a linear disposable income, allowing straightforward economic welfare analysis (Williams 1977). Income interactions were also included with the zonal average income to identify tendencies for income clustering, and with the square footage of housing units. A size term is included to account for varying numbers of units within a building – since the date include single-family buildings, condominiums, flats, and apartments. This specification is not final, but reflects an example of how dwelling level attributes (price, square footage, lot size), zone level attributes (average income, accessibility to employment by different modes, density), and household attributes (income, number of workers) are reflected in the specification of the model. It is straightforward to add other variables that draw on characteristics of the built, social, and economic environment, using simple expression syntax to define new variables.

These estimation results, while not final, reflect reasonable results and significance. The disposable income variable (income – annual rent) was positive and significant, which is noteworthy since it is not uncommon in discrete choice models of housing location to find insignificant or even counter-intuitive signs on price variables, due to omitted variables that are correlated with price. The goodness of fit is also relatively high for disaggregate, household-level discrete choice models. It is interesting to note that for households with lower numbers of workers, the bus access measures dominated the auto measure in the San Francisco area. This is not perhaps very generalizable to other cities in the U.S. but provides some evidence of the influence of transit access on residential choices in places with high levels of transit service.

6 Calibration and Validation

UrbanSim choice models such as the household location choice models involve very large numbers of alternatives. In the Paris application, there were 1,300 Communes used, and it was possible to enumerate all of them in a choice model. But generally, it is necessary to sample alternatives rather than enumerate the entire universal choice set. In the Puget Sound application, there are approximately 1.2 million parcels, for example. With choice models that use random sampling of alternatives, there are no alternative specific constants being estimated, which might require calibration in the way that mode choice models, by contrast, typically require to match aggregate mode shares.

While there are no alternative-specific constants to calibrate in a spatially detailed location choice model, it is still possible to include dummy variables reflecting larger districts, to capture unobserved characteristics of areas that might otherwise bias the predictions in those areas. This is not recommended, in general, since including such constants excessively can constrain the model and make it less policy-sensitive, and it is not clear whether or how such constants should change in the future.

Since UrbanSim is a stochastic microsimulation model, which means there is random variation arising from the use of random draws to make choices from probability distributions. Some have raised questions about this simulation "error" potentially being quite large. In addition there is uncertainty arising from

Table 3 Results from calibration using Bayesian melding in Eugene-Springfield Model application

Method	Number of cases missed out of 265	Coverage by 90% confidence interval
Multiple runs only	163	0.38
Bayesian Melding	31	0.88

errors in the input data, uncertainty in the model parameter estimates, and even uncertainty in the model structure. It is important to develop ways of handling this uncertainty in a principled way, and calibrating the uncertainty in the model in ways that provide more robust capacity to make policy assessments.

Sevcikova et al. (2007) have adapted a Bayesian Melding technique for calibrating stochastic simulation models such as UrbanSim. The technique was originally developed by Raftery et al. (1995) for use in deterministic models. The process requires longitudinal data, but provides a way to rigorously calibrate the uncertainty in the model system in order to make statistically valid inferences regarding the results. Due to random variation from the stochastic nature of the model, and other sources of uncertainty such as input data and parameters, running the model system multiple times generates a distribution of results. In a properly calibrated model, run over a period of years, the 90% confidence interval computed from the distributions of the results should cover the observed data 90% of the time. Our application of this technique to the Eugene-Springfield model application shows that if we only account for the random variation in the simulation, the 90% confidence interval from the results of 15 years of simulation only covers the observed outcomes in the real world 38% of the time, as shown in Table 3. After using the Bayesian Melding to calibrate the model system, this coverage increased to 88%, which reflects a well-calibrated result, using households and employment by traffic analysis zone as the basis for the calibration. This process of calibration is computationally expensive, requiring multiple runs of the model system. Our results achieved a high degree of convergence within approximately 200 runs, and experiments using much larger numbers of runs (3,000) did not significantly change the results.

The Bayesian Melding technique has now been applied to the Eugene-Springfield application, and work is underway to apply it to the Puget Sound model application, to support analyses such as the comparison of alternatives for replacing the earthquake-damaged Alaskan Way Viaduct. Considering the controversy surrounding this project, it is certainly not clear that more informed modelling will influence the outcome of the political process, but it is worth investigating whether it could move the debates to a more productive focus.

7 Software Implementation

UrbanSim is currently implemented in the Python programming language, using an Open Source (GPL) license. It is available for download from the project website at www.urbansim.org. The software platform is called the Open Platform for Urban

Simulation (OPUS). The decision to convert UrbanSim from Java to Python, and to develop the OPUS platform, grew out of interactions with research groups developing transportation and land use models in North America, Europe and Asia, all of whom needed to develop their own software applications, and found that they were spending far too much time on developing and debugging software, and far too little on developing models, applications and new research. An initiative emerged following a meeting in Toronto in January 2005, to develop an open platform that could be shared among researchers and practitioners for land use and transportation model development, allowing people to more easily leverage the work of others and to make their own investments more effective. The OPUS architecture is intended to facilitate collaborative development, and contributions of packages by a community of users and developers.

The UrbanSim development team has led the development of OPUS, and has ported UrbanSim to it. This effort was completed in 2006, and a new release in 2008 added a flexible Graphical User Interface for creating models, estimating their parameters, and combining models into model systems that are run on policy scenarios. An international working group has been established to further develop and refine OPUS, and to begin to provide a stable, shared laboratory for collaboration, and for rapid development, testing and comparison of alternative algorithms and models.

The OPUS Architecture is three-tiered, with the Opus Core forming the foundation, a set of OPUS Packages extending this, and a set of external libraries that provide access to functionality in external systems and languages. OPUS packages are all implanted in Python, but external libraries may be in C, C++, or Fortran. Interfaces to a range of databases and flat files are available, including MySQL, MS SQL Server, Postgres (and PostGIS), SQLite, DBF, CSV, and Tab-delimited ASCII files. Access to data in ArcGIS is integrated, as well as interaction with the open source PostGIS system.

8 Conclusions

The household location choice model of UrbanSim has evolved over multiple applications in the United States and Europe to account for variation in social and political context as well as in data availability. It has been applied in places as small as Washtenaw County Michigan, and as large as the Paris Metropolitan area in France, with 11 million inhabitants. The unit of analysis for location choice has varied among applications from gridcells to parcels to zones and in Paris, Communes. UrbanSim and the Open Platform for Urban Simulation have, as a result of the needs of these applications, been redesigned to provide a highly modular, flexible framework to support rapid experimentation and development of new models and approaches, along with the computational performance needed for large-scale production use.

As an example of a parcel-level model, results cited in this paper from the San Francisco model application show that the estimation produces quite significant and sensible results, and that the goodness-of-fit is quite good for a disaggregate choice model. Similarly, we found that run times were reasonable, with the Household Location Choice Model running at the parcel and building level for all of San Francisco in 5 minutes on a standard desktop computer, and the entire model system running in approximately one to two minutes per simulated year. The total development effort for the San Francisco model application, including data compilation and processing, model specification, creation of the parcel version of the models, and estimation of the model parameters, was approximately person-months of effort over 18 months. This reflects significant progress over earlier applications of UrbanSim, which have generally required substantially greater effort to develop. This efficiency was due to a combination of excellent data available from the San Francisco County Transportation Authority and the City of San Francisco Planning Department, and the modularity of OPUS and UrbanSim allowing rapid development of new parcel-level models.

As always, much more remains to be done. Recent research has focused on improving the capacity to estimate flexible models with interdependent, non-nested choice dimensions such as residence and workplace (Waddell et al. 2007b). Work is underway to develop a Graphical User Interface, to improve integration with GIS, and to develop more documentation and tutorials. Data mining and imputation methods to ease the task of creating the data needed for the model are being investigated. Current research projects focus on the integration of activity-based travel modelling and dynamic traffic assignment, and on the evaluation of complex land use and transportation policy scenarios with regard to their impacts on travel behavior, urban form, emissions and air quality.

Acknowledgements This research has been funded in part by National Science Foundation Grants IIS-0705898 and IIS-0534094, by EPA Grant R831837, and by the San Francisco County Transportation Authority. OPUS and UrbanSim have been developed by a team of collaborators, coordinated by the author. Particular acknowledgment is due to Alan Borning, Hana Sevcikova and Liming Wang, along with many other collaborators and contributors.

References

Beckman R-J, Baggerly K-A, McKay M-D (1996) Creating synthetic baseline populations. Transp Res A Policy Pract 30(6):415–429

Ben-Akiva M, Lerman S-R (1985) Discrete choice analysis: theory and application to travel demand. MIT, Cambridge

de Palma A, Picard NA, Waddell P (2007) Discrete choice models with capacity constraints: an empirical analysis of the housing market of the greater Paris region. J Urban Econ 62:204–230

Lee D (1973) Requiem for large scale urban models. J Am Inst Plann 39(3):163–178

McFadden D (1973) Conditional logit analysis of qualitative choice behaviour. In: Zarembka P (ed) Frontiers in econometrics. Academic Press, New York, pp 105–142

Raftery AE, Givens GH, Zeh JE (1995) Inference from a deterministic population dynamics model for bowhead whales. J Am Stat Assoc 90(430):402–416

Sevcikova H, Raftery A, Waddell P (2007) Assessing uncertainty in urban simulations using Bayesian melding. Transp Res B Methodol 41(6):652–659

Train K (2003) Discrete choice methods with simulation. Cambridge University Press, Cambridge

Waddell P (1998) The Oregon prototype metropolitan land use model. In: Easa S, Samdahl D (eds) Transportation, land use and air quality: making the connection. American Society of Civil Engineers, Reston, VA, pp 549–558

Waddell P (2000) A behavioral simulation model for metropolitan policy analysis and planning: residential location and housing market components of UrbanSim. Environ Plann B Plann Des 27:247–263

Waddell P (2002) UrbanSim: modeling urban development for land use, transportation and environmental planning. J Am Plann Assoc 68(3):297–314

Waddell P, Borning A, Noth M, Freier N, Becke M, Ulfarsson G-F (2003) Microsimulation of urban development and location choices: design and implementation of UrbanSim. Netw Spat Econ 3(1):43–67

Waddell P, Ulfarsson G-F, Franklin J, Lobb J (2007a) Incorporating land use in metropolitan transportation planning. Transp Res A Policy Pract 41:382–410

Waddell P, Bhat C, Eluruu N, Wang L, Pendyala R (2007b) Modeling the interdependence in household residence and workplace choices. Transp Res Rec 2003:84–92

Williams HCWL (1977) On the formation of travel demand models and economic evaluation measures of user benefit. Environ Plann A 9:285–344

Household Behaviour in the Oregon2 Model

J.D. Hunt, J.E. Abraham and T.J. Weidner

Abstract The Oregon2 Model represents the spatial activity system in the State of Oregon in the United States. It uses a set of seven connected modules representing different components of the full system, each running in turn for each year of simulation. Two of the module concern elements of household behaviour. The Household Allocations Module provides an agent-based microsimulation of each household and each person, simulating the transitions and choices made by these agents over 1 year. The Land Development Module provides a representation of space development using 30 m x 30 m grid cells covering the model area, micro-simulating development transitions occurring in each cell over 1 year. It determines changes in developed space over time and in response to potential policy actions involving pricing, regulation and infrastructure in both transportation and land use. At the time of writing, the system is not yet complete in that much of the second and all of the third stages of calibration are still outstanding. But some preliminary conclusions about the design and development of the model system, and the treatment of households in particular, can still be offered.

J.D. Hunt (✉)
Department of Civil Engineering, University of Calgary, 2500 University Drive NW, Calgary, AB, T2N 2N4
e-mail: dhunt@ucalgary.ca

J.E. Abraham
HBA Specto Incorporated, 101 – 6th Avenue SW, Suite 1060, Calgary, AB, T2P 3P4, Canada
e-mail: jea@hbaspecto.com

T.J. Weidner
Parsons Brinckerhoff Inc, 400 SW Sixth Avenue, Suite 802, Portland, OR, 97204, USA
e-mail: weidner@pbworld.com

F. Pagliara et al. (eds.), *Residential Location Choice*, Advances in Spatial Science,
DOI 10.1007/978-3-642-12788-5_9, © Springer-Verlag Berlin Heidelberg 2010

1 Introduction

This chapter describes the representation of household behaviour in the Oregon2 Model. The first section is this introduction. The second section overviews the Oregon2 Model framework. The third section presents the Household Allocations (HA) Module, which includes a microsimulation treatment of household and person-level demographic transitions and relevant location decisions. The fourth section covers the Land Development (LD) Module, which includes a disaggregate treatment of residential floorspace development related to household behaviour along with other types of floorspace. The fifth section discusses issues in parameter estimation and model calibration. The sixth and final section offers some conclusions.

2 Oregon2 Model Framework

The Oregon2 Model represents the spatial activity system in the State of Oregon using a set of seven connected modules that cover different components of the full system. This set of modules and the flows of information among them are shown in Fig. 1. Each module is run in turn for each year of simulation, starting at the top (ED) and working around the circle in a clockwise pattern, with the results of

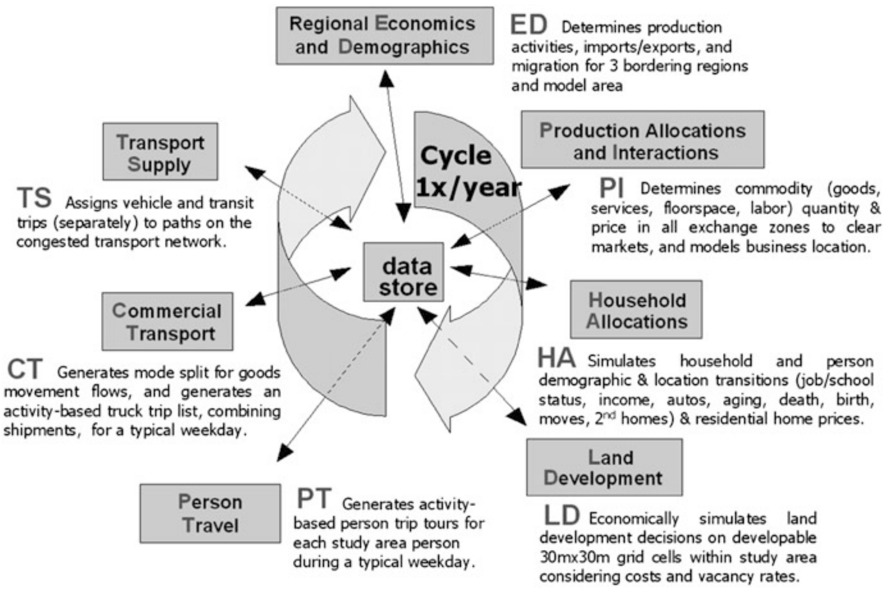


Fig. 1 Oregon2 model system. For each year simulated, each module is run in turn starting with ED and proceeding clockwise. The results from each module are sent to the data store and thus are available for subsequent module runs

particular modules influencing those of other modules run subsequently in the same year or subsequent years.

The ED and PI Modules use aggregate, equilibrium approaches to identify solution states. The others use disaggregate, microsimulation approaches to mimic real-world processes – with the exception of TS, which first identifies an aggregate, equilibrium solution and then considers each vehicle and person trip while updating the representation of aggregate conditions to reflect changes arising with this consideration.

Work on the development of the Oregon2 Model is ongoing at the time of writing. A working version of each module has been established, and the calibration of these modules continues, first considering each module separately and then considering the full system of modules working as a unit.

An Oregon2 Transitional Model has also been developed by substituting more aggregate treatments of the household behaviour and space development process (HA and LD Modules) and the assignment process (TS Module). The Transitional Model focuses the modelling team’s efforts on calibrating and refining a simpler model that can address pressing state policy questions that needed to be analyzed before the full Oregon2 model could be calibrated. It is anticipated that further advancements on the HA and LD Modules described here will continue once the Oregon2 Transitional Model is calibrated and under application.

3 Household Allocations Module

The Household Allocations (HA) Module provides a fully disaggregate representation using an agent-based microsimulation of each household and each person, simulating the transitions and choices made by these agents over the period of 1 year. The intent is to perform an endogenous determination of changes in social characteristics, so as to provide a more complete and consistent representation of demographic changes over time and in response to a wide range of potential policy actions involving pricing, regulation and infrastructure in both transportation and land use.

3.1 Definitions and Categories

Table 1 shows the attributes of individual households and household members tracked in the microsimulation, together with the coding for each. In most cases the coding vectors are those used or adapted from the United States Census (US Census 1990a).

Locations in space are represented using the system of zones shown in Fig. 2. These zones are called “alpha zones” in order to distinguish them from other types of zones used in the Oregon2 Model. A total of 2,951 of these alpha zones cover the 36 Oregon Counties plus 39 halo counties in adjacent states. They vary in size in rough proportion to population, and about half are in urbanized areas. Those inside

Table 1 Household and person attributes in HA module. Each household and its members are described in terms of these attributes, and the states assigned to these attributes can be changed as part of the consideration of the household and its members in the model

Household attributes		Person attributes	
Code	Description	Code	Description
HH_ID	Household ID	HH_ID	Household ID
TAZ	Alpha Zone 0001-4141	PER_ID	Person ID (HH_ID*100 + PER_ID)
XGRID	GridCell Matrix X-Position XXXXX. X-Position	AGE	00. Less than 1 year 01. 89. Age in years 90. 90 or more years old
YGRID	GridCell Matrix Y-Position YYYYY. Y-Position	SEX	0. Male 1. Female
RHHINC	Household income (1990$) 0000000. N/A (GQ/vacant/no income) −999999.9999999. Total household income	RLABOR	Employment status 01. Employed 06. Not in labor force
AUTOS	Vehicles (1 ton or less) available 0. N/A (GQ/vacant) 1. No vehicles 2–7. 1–6 vehicles 8. 7 or more vehicles	HOURS	Hours worked last week 00. Not employed 20. Part time (not currently used) 40. Full time
YRMOVED	When moved into this house or apartment 00. Moved in current year x>=1. Moved x years ago, relative to current year (add 1 to last year, if not moved)	FERTIL	Number of children ever born 00. N/A (less than 10 years/ male) 01. No children >01. Number of children born plus 1
UNITS1	Units in Structure: 01. Mobile home (MH) – OR – Rural Residential Mobile Home (RRMH) 02. One-family house detached (SFD) –OR – Rural Residential single detached (RRSF) 03. Two-family house attached (AT) 05. Apartments (MF)	YEARSCH	Years of Educational attainment (recoded) 00. No primary school completed 01–10. 1st –10th grade 11. 11th –12th grade, no diploma 12. High school grad, diploma / GED 13–15. Some college, but no degree 16–17. Associate, Prof., BSc degree 18–20. MSc degree 21. PhD / DSc degree
ONEACRE	House on 1 acre of land or more: 2. Rural Residential (RRSF or RRMH) 1. Otherwise (SFD, MH, AT, MF)	OCCUP	000. Not Employed 001. Mangrs&Professionals (1_ManPro) 083. Health workers (1A_Health) 113. Post secondary teachers (2_PstSec) 155. Non-P.S. teachers (3_OthTchr) 163. Other Prof. & Tech. Ofc (4_OthP&T) 263. Retail sales workers (5_RetSls) 283. Other Retail&Clerical Ofc (6_OthR&C) 500. All other (7_NonOfc)
VAC	0000. No Secondary Home		
HOMETAZ	0001-4141. Alpha Zone of Secondary Home		
		WORKTAZ	0001-4141. Alpha Zone of Work location

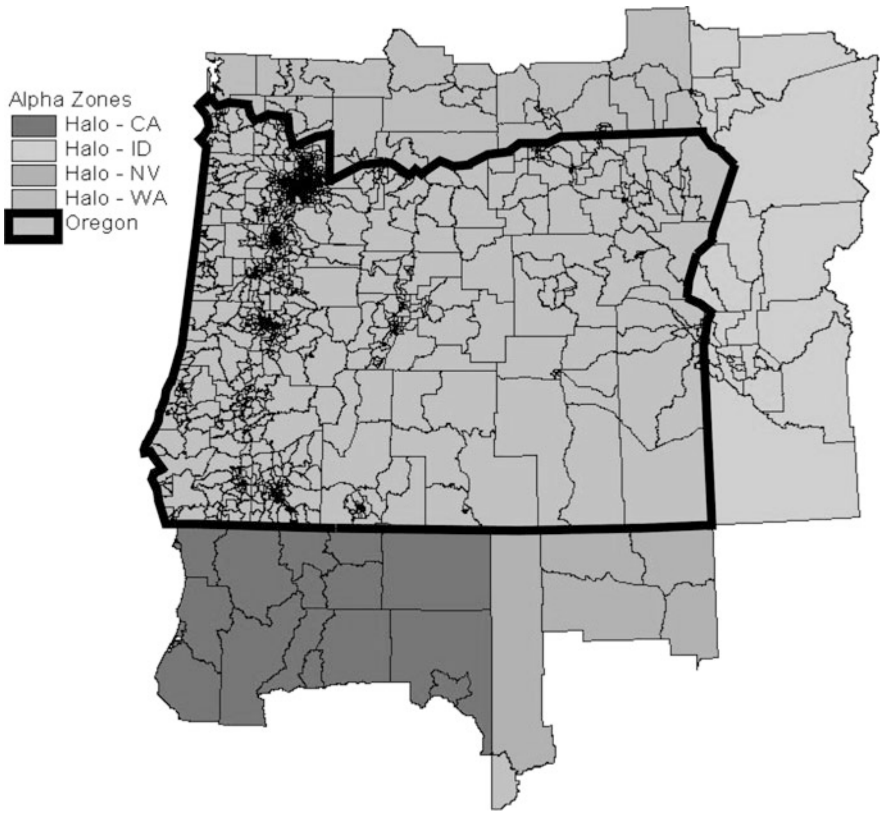


Fig. 2 Oregon2 alpha zone system. A total of 2,951 geographic zones cover the model area, which includes all of Oregon and a "halo" in adjacent states. These zones are called "alpha zones" to distinguish them from other types of zones used in the Oregon2 Model

Metropolitan Planning Organizations (MPOs) match the boundaries used in the corresponding local transportation models; those outside MPOs nest within US Census block groups.

The model area is also represented using a system of almost a billion 30 m x 30 m grid cells, each of which contains no more than one category of developed space, with specific type, age and development intensity, and land, with specific zoning regulations.

The HA Module uses the subset of these grid cells with residential space as an inventory of the supply of residential space, and treats this as a fixed supply within a given year. The LD Module, described below, updates the supply of residential space between runs of the HA Module, partly in response to the space prices (rents) established in the HA Module. The LD Module similarly updates the supply of non-residential space in response to non-residential space prices (rents) established in the PI Module.

An initial population of households and household members for use in the simulation, with all attribute values assigned, is synthesized for the year 1990 using a sampling process that draws on a disaggregate sample of actual households and relevant marginal distributions from the US Census (Beckman et al. 1996; US Census 1990a, b).

3.2 *Process*

In a given run of the module, for a particular year, each household is considered in turn in a random order. When a household is considered, a series of special sub-models is used to take the household and its members through possible transitions and update their attributes accordingly. This series of sub-models is shown in Fig. 3.

Some of these sub-models are rule-based. For example, each person's age increases by 1 each year. Most are probabilistic, with Monte Carlo techniques used to assign states to attribute variables. The selection probabilities are assigned to possible states based on the attributes of the household and its members and the states themselves, in most cases using some form of single-level or nested logit choice model. The use of logit formulations in this way permits the further representation of changes in user benefits through the calculation of the log-sum (also called the "inclusive value" or the "expected maximum utility") over the range of available alternatives (Williams 1977).

Changes in household and person status, and the possibility of a "non home anchored job search", are considered first, followed by changes in residential location regarding both primary and secondary homes. Then the interactions among households are processed, followed by auto ownership. Finally, the allocation of jobs is considered.

Within a particular year interactions occur between households in both (a) the residential real estate market and (b) the movement of persons out of their existing

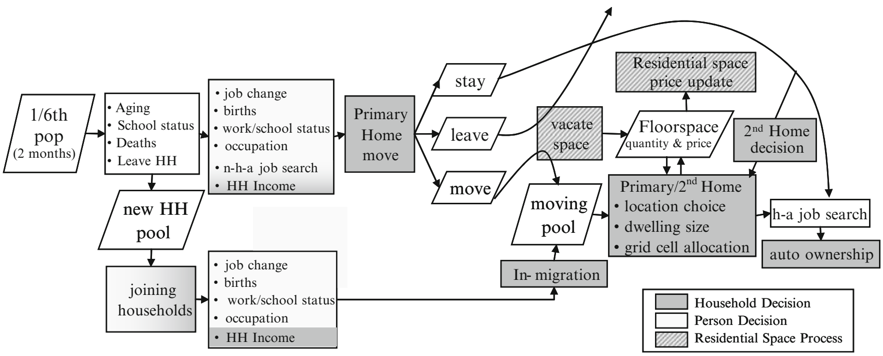


Fig. 3 HA module process diagram. For each year simulated, each household and its members are taken through this process covering demographic transitions, household formation, location and auto ownership decisions and person job and occupation decisions. Use of residential space and changes in household membership and associated formation of new households are also tracked

households and into new households (possibly single person households). These interactions are resolved by accounting for the departure of households from existing space and the departure of persons from existing households over a fraction of the full year, and at the end of this fraction moving households are allocated back into the available vacated existing space and departing persons into new households. This is done over a fraction of the full year in order to provide a more accurate representation of the dwellings available to choose from in the residential real estate market, and of the speed of price responses in the real estate market. Currently, six equal-sized fractions are used per year – where 1/6th of the population is processed in each such fraction – notionally consistent with a 2-month lag between one household listing a residential space for sale or lease and another household subsequently occupying the given residence. The use of a fraction other than 1/6th potentially could be considered as part of the later calibration of the model.

Interactions between members of the same household are accounted for within the individual submodels. One such set of interactions considered explicitly each year concerns the home location for a household and the job locations for its members. These interactions are shown diagrammatically in Fig. 4.

Initially, there are a certain number of people with jobs in the prior year (black circle). Some of these people become unemployed (the area labeled "became unemployed") and thus lose both their job and their job location. Other people previously not employed become employed (the area labeled "became employed"). As well, some of those people who were employed and remain employed will undertake a job search (the area labeled "left job location"). At this point there are a number of people who are to be employed but do not have job locations. Some of these people are selected to undertake a non-home-anchored job choice

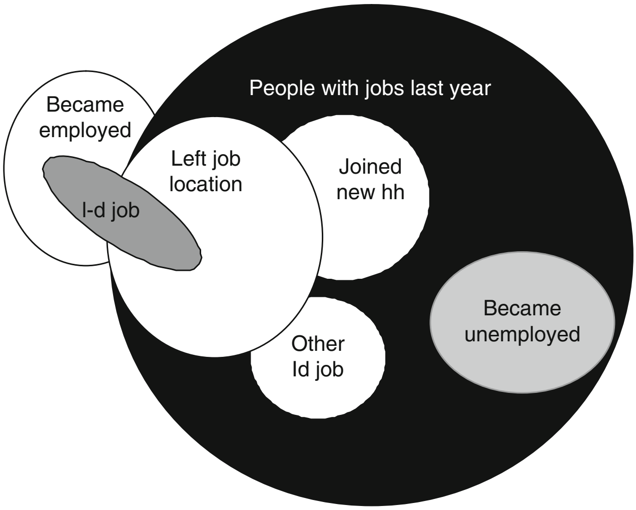


Fig. 4 Home and job location interactions. This Venn diagram shows the treatments applied to the sub-sets of persons with different job and home location situations arising in the simulation

(ellipse labeled "ld job" for "long distance job"). These people then acquire new jobs without regard to their current home location.

The households that have undertaken the long distance job search may contain more than one employed household member whose jobs may now be unsuitable, as there may be no possible home location within reasonable commute distance to the set of household jobs.

To correct this, the other people in the household are flagged to do a home anchored job search from the new household location (the area labeled "other ld job" for "other long distance job").

New households are created by joining up those people who left existing households. These people may already be employed with jobs in locations that are too far apart for there to be a suitable home location. To correct this, just one of the employed persons in each newly joined household retains the current job and corresponding job location, and the other employed people in the household are flagged to do a home anchored job search (the area labeled "joined new hh").

At this point, households with more than one employed household member may have more than one assigned job location, but only if all of those job locations existed from the previous year. Newly created households, or households that have undertaken a long distance job search, will have no more than one assigned job location, with other household workers flagged for a home-based job search.

This set of assigned job locations then serves to inform the home location choices. A household is more likely to move if its home is a long way away from this set of jobs, and a household is more likely, all other things being equal, to move to a new home location close to this set of jobs.

The four white areas in Fig. 4 consist of employed people who do not have job locations assigned to them from the previous time period or from the non-home-anchored job choice model. These people will perform a home anchored job search at the end of the process, after their households have had a chance to consider their home locations and, possibly, move to other locations in the model area.

3.3 Primary Home Choice Structure

The entire choice model concerning primary homes is incorporated in a combination of the sub-models; the overall structure of this entire choice model is shown in Fig. 5.

In Fig. 5, solid lines indicate two-way relationships, where the lower level choices are conditional on the higher level choices and where the log-sum values for the lower level choices feed back up to influence the higher level choices. Dashed lines indicate one-directional relationships, where the lower level choices are conditional upon the higher level choices, but the higher level choices are not influenced by the log-sum values for the lower level choices.

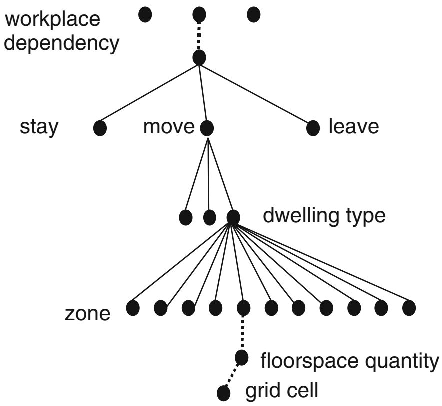


Fig. 5 Primary home choice structure. The combination of sub-models considering the aspects of home choice results in a set of conditional choice models, some (non-dashed) of which are linked in a nested logit structure

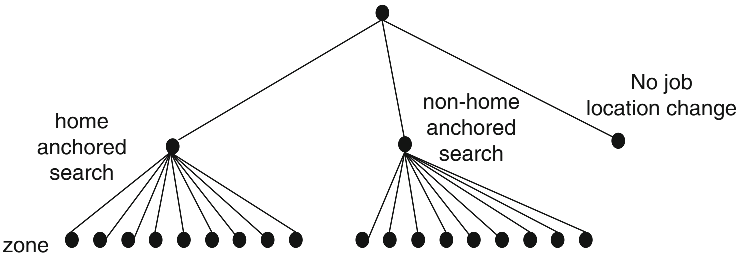


Fig. 6 Job choice model nesting structure. The probabilities that individuals move to new jobs, using either home-anchored or a non-home-anchored search processes, are influenced by the log-sum (composite) utilities for the relevant sets of job alternatives

3.4 Job Choice Structure

The flow of conditional choices and the calculation of composite utilities are such that the entire job choice model has a nested logit structure as shown in Fig. 6.

Job location changes are done at the person level, with the type of job change (top level of nesting structure) and the non-home-anchored searches being done before the home choices, and the home-anchored searches being done after home choices. Thus the alternatives in the home-anchored search could have different utility values when they are being evaluated to calculate the log-sum for the higher level of nesting than the values they will have later if they are evaluated again to select a particular job location.

Considering the HA sub-models in detail, as shown in Fig. 3:

In-migration. This is applied once at the start of each current year. It creates the list of households in-migrating to the model area in that year, by randomly (with replacement) selecting households from an in-migrating subset of the 1990 PUMS sample of households, including all those households that moved to Oregon between 1985 and 1990. The number of in-migrating employed people by industry and the number of in-migrating retired and unemployed households are targets that are reached by selecting enough households of appropriate types. Each selected household is added to the primary home moving pool as long as adding it to the population does not cause any targets to be exceeded. Households continue to be selected until all the targets are reached.

Aging. This is applied to each person and each household. The age of each person is increased by one. For those people who are students, the number of years at school is also increased by one. The number of years at the current home location is increased by one for each household.

Deaths. This is applied to each person, selecting between the live and die possibilities for the current model period. Age and gender influence the probability that a given person dies. A table of annual death probabilities for different age and gender categories from Oregon Vital Statistics Reports (Oregon Department of Human Services 1998) is used for the Monte Carlo process.

Births. This is applied to each woman with age between 10 and 49 years inclusive. It determines the resulting number of babies born, if any, and their genders through a four-step Monte Carlo process using fixed probabilities. In the first step each woman is assigned a married / non-married status, where the probability of marriage is a fixed non-zero value when there is at least one male of age in the woman's household and zero otherwise. In the second step each woman is assigned a "give birth"/"not give birth" status for this time period, based on her age, number of previous babies, and marriage status as assigned in the first step. In the third step each woman assigned the "give birth" status in the second step is assigned a number of babies for this birth, drawing from "1", "2" and "3" based on the woman's age. In the fourth step each baby born is assigned a gender, with a fixed probability of the baby being male. The probabilities at each step are based on National and Oregon Vital Statistics Reports (National Center for Health Statistics 1999, 2000; Oregon Department of Human Services 1995).

Work and school status. This is applied to each person in each time period. It assigns the person a work status (either "worker" or "non-worker") and a school status (either "student" or "non-student") for the time period in a single joint process. A logit model is used to assign the probabilities to the alternatives for the Monte Carlo process, with the utility functions for the alternatives including the person's age, gender, education level, previous work (and occupation) and school status, household size, presence of children in the household, and log-sum calculated from the home anchored job search model, based on data from the Panel Survey of Income Dynamics (PSID) (University of Michigan 1981–1993).

Occupation. There are two versions of the occupation choice sub-model. One is the "transition occupation choice" version, which is applied to each person who is currently assigned an occupation to determine if there is a change in occupation. The other is the "open occupation choice" version, which is applied to each person who is not currently assigned an occupation. The "open occupation choice" version is used the same year a person switches from "non-worker" to "worker" status by the work and school status sub-model, regardless of whether or not the person was, in some year prior to the previous year, assigned "worker" status. Both versions assign each person an occupation status from the eight categories listed in Table 1. The factors influencing the probability that a given status is assigned include the person's age, gender, education level, current occupation (for the "transition occupation choice" version), presence of children in household and household size, and utilities indicating the labor market conditions for the various occupations near the zone containing the person's household as determined in the PI Module. In both versions, a logit model is used to assign probabilities to the occupation status alternatives for the Monte Carlo process. The utility functions for the "open occupation choice" version do not include constants for transition between occupation status alternatives, otherwise the same alternative specific constants are used in both versions. Parameters are estimated using data from both the US Census (US Census 1990b) and the PSID (University of Michigan 1981–1993).

Leave household. This is applied to each person in a household with more than one member. The person is assigned to either "leave" or "stay" in the current household. A Monte Carlo process is used with the selection probabilities determined using a logit model. The utility functions for the alternatives include the person's age, level of educational attainment, work status, person school status presence of children in household and household size, with parameters estimating using the PSID data (University of Michigan 1981–1993). Any person assigned to "leave" is then added to the join household pool, and thus is considered in the household joining sub-model.

Joining households. This is applied to the set of persons currently in the join household pool in each time period. It places persons from this pool into new households, thereby removing them from the pool, continuing until the pool is empty. The resulting new households are then added to the primary home moving pool, and thus considered in the primary home location sub-model. Currently, the model does not allow individuals to join existing established households formed initially in prior years, but rather only joins together persons who left other households thereby always forming new households of types consistent with observations. The process starts by forming a random order list of the persons currently in the join household pool. Then, a household category is selected randomly from an exogenously specified set of new household categories, and with specified selection probabilities. These household categories indicate the composition of candidate new households in terms of numbers of members, their genders and age ranges. Starting at the top of the random order list of persons currently in the join household pool and working down the list in order, each person is checked to see if he or she fits within the candidate new household. If the person does fit, then he or she is placed in the candidate new household and the process continues down the person list seeking to fill the remaining places in the candidate new household. Once all the places in the candidate new household have been filled, then the persons placed in it are removed from the person list and the candidate new household becomes a new household. The process then selects another household category and starts again at the top of the person list seeking to fill the places in the next candidate new household. If the end of the person list is reached before all the places in the candidate new household have been filled, then the candidate new household is no longer considered and all of the persons that have been placed in it so far are not removed from the person list. The process then selects another household category and starts again at the top of the person list. This continues until the person list is empty. In order for the process to be able to finish, the set of new household categories must include a sub-set of categories that spans all the possible forms of one-person households. Otherwise, the process could get caught in an infinite loop where only one person remains in the person list.

Job change. This is applied to each person currently assigned "worker" status by the work and school status sub-model, implying that the person has a current job. It determines if the person changes jobs, and in some cases also updates the jobs holdings for the other members of the person's household, as discussed above (Fig. 4). Each person is assigned to one of the following categories regarding job holdings: "keep current job", "switch to new job with a home-anchored search" or "switch to new job with a non-home anchored search". This is done using a logit model with utility functions that include age, occupation, household composition, as well as log-sums from the job location choice model. When a person is assigned to the "switch to new job with home-anchored search" category, this person is placed in the home-anchored job search pool, and thus is considered in the home-anchored job search sub-model. When a person is assigned to the "switches to new job with non-home-anchored search" category, then this person is placed in the non-home-anchored job search pool, and thus is considered in the non-home-anchored job search sub-model. In this case the person is considering new jobs that are beyond reasonable commuting distance from the current primary home location. As discussed above, other members of the household who are currently assigned "worker" status by the worker and school status sub-model are assigned to, "switch to new job with home-anchored search", and are considered in the home-anchored job search sub-model after the new home location has been identified. The data used in estimation were drawn from the US Census (US Census 1990b) and the PSID (University of Michigan 1981–1993). The factors influencing the probability that a given job holdings category is assigned include the person's age, occupation, household composition, and composite utility functions (log-sums) from the home anchored job location choice sub-model and the non-home-anchored job location choice sub-model. Probabilities for the Monte Carlo process are assigned to the alternative categories using a logit model. The utility functions for the two "switch to new job" alternatives include composite utilities for the jobs over the full set of zones. This places the job change, the home-anchored job search and the non-home-anchored job search sub-models within a formal nested logit structure as indicated in Fig. 5 and discussed above.

Home-anchored (h-a) job search. This is applied to each person in the home-anchored job search pool. It assigns a new job to the person, specifically assigning an alpha zone that contains the location of the job. The factors influencing the probability that a given alpha zone is assigned are the comparative ease of traveling from the home location, the distribution of wage rates and the relative size of employment for the person's occupation category in each alpha zone. A logit model is used to assign probabilities to these alternatives for the Monte Carlo process.

Non-home-anchored (n-h-a) job search. This is applied to each person in the non-home-anchored job search pool. It assigns a new job location to the person, specifically assigning an alpha zone that contains the location of the job. The probabilities for the alpha zones used in the Monte Carlo process are based on a logit model, with the same utility function for each zone as in the home-anchored job search above, excluding the ease of travel from the current home location.

Household income. This is applied to each household each year. It assigns a total annual before tax income to the household. The factors influencing the assigned value include household attributes (workers by occupation, students, pre-school children, unemployed, retired) and the unit prices for labor for the relevant occupations in the zones containing the workplace locations for each worker in the household. In this process, those household members who are both "student" and "worker" are treated according to just their worker attributes; and those who are both "non-student" and "non-worker" are designated either "unemployed" or "retired" using a Monte Carlo process with probabilities that vary by age according to the Census PUMS data (US Census 1990a). Then, the total household income is determined by summing the contributions made by the members of the household according to their status in these categories, with a randomly varying amount also added.

Primary home move. This is applied to each household currently assigned a primary home location zone and dwelling type. It determines whether the household "moves" or "stays" (remains in the same dwelling) and if it "moves", whether or not it moves out of the study area. A Monte Carlo process is used, with the selection probabilities determined using a nested logit model. Household composition, accessibilities to current workplaces, distance to previous home location, distribution of vacant residential space, and residential space prices all influence the utility values for the alternatives. More specifically, the utility functions for the "stays" and "moves" alternatives are consistent with the corresponding utility functions for the primary home location sub-model described below. The same values are used for the matching parameters. The utility function for the "stays" category is the utility function for a specific location choice alternative (the location currently assigned the household), with the zonal size term removed. The utility function for the "moves" category is the composite utility for the full set of location choice alternatives. This places the primary home move and primary home location sub-models together within a consistent nested logit structure. When a household is assigned the "moves" category, it is added to the primary home moving pool, and thus is considered in the primary home location sub-model.

Primary home location. This is applied to each household in the primary home moving pool. It jointly assigns the household primary home an alpha zone that contains the home location and a dwelling type from among the six categories listed in Table 1. The factors influencing this joint assignment are the same as those acting in the primary home move sub-model. A Monte Carlo process is used to select the alpha zone and dwelling type, with the selection probabilities for the joint alternatives determined using a nested logit model where the components of utility of a given dwelling type in a given zone include a representation of ease of travel to assigned workplaces (as a log-sum over the available modes of transport to work, calculated by the PT module), the distance to the previous home location (if any), the quantities of vacant residential space of the dwelling type in the zone (a size term), and the residential space prices. These are in a nested logit model, so that alternatives that share a common dwelling type can be more similar (i.e. have a smaller expected size of the stochastic term of the utility) than alternatives that share a common zone, and households will be more willing to change zones than dwelling type with a calibrated parameter. The logit model was estimated using data from the Oregon Travel Behavior Survey (Oregon Department of Transportation 1994, 1996).

Primary home dwelling size. This is applied to each primary home assigned a location zone and dwelling type in the primary home location choice sub-model in that year. It determines the resulting number of rooms and quantity of residential space for the primary home using a two-step process. In the first step, the home is assigned a number of rooms (as a continuous value, not as an integer) using a linear equation where the independent variables are the dwelling type, the number of workers in the household, the household income, the number of children under 18-years old in the household and the residential space price in the location zone. In the second step, the home is assigned a quantity of residential floorspace (in square feet) according to the number of rooms value assigned in the first step and using a function that performs a linear interpolation between specified quantities of space for integer numbers of rooms, with these quantities varying by dwelling type. The data used to estimate the equations used in these two steps were drawn from the US Census PUMS (US Census 1990b) and the American Housing Survey (US Census 2002).

Secondary home decision. This is applied to each household in the study area, assigning the household a secondary (or vacation) home status from the “has second home” and “does not have second home” alternatives. This is done using a Monte Carlo process with a binary logit model assigning the selection probabilities to the two alternatives. One of the attributes of the utility function for the “has second home” alternative is whether the household had a secondary home in the previous year – establishing the representation that households with secondary homes in 1 year are more likely to have them in the next year. For a household that had a secondary home in the past year and is assigned to have a secondary home in the current year, there is a further determination regarding whether the household retains the current secondary home or vacates it and obtains a new one. This is done using a Monte Carlo process with fixed probabilities for the two possibilities. Whenever a household is determined to vacate a secondary home (regardless of whether or not to obtain a new one) the floorspace for the vacated home is added to the floorspace inventory at the end of the sub-timestep. Those households determined to obtain new secondary homes are added to the secondary home moving pool, so they are considered in the secondary home location sub-model. At the time of writing, the utility functions for the “has second home” and “does not have second home” alternatives include only a few variables related to characteristics of the household; an expansion to add further representation of the influences of other relevant factors is part of planned future improvements to the model regarding the treatment of secondary homes.

Secondary home location. This is applied to each household in the new secondary home pool. It assigns the household secondary home to an alpha zone. A Monte Carlo process is used to select the zone, with the selection probabilities for each alternative zone determined using a utility function that includes the price of available floorspace in the zone, the distance from the zone to the household primary home location, and a zonal constant reflecting the quantity of vacation homes available in the zone. Currently, only single-family dwelling types are used for secondary homes in the model; an expansion to consideration of other dwelling types is part of planned future improvements to the model regarding the treatment of secondary homes.

Secondary home dwelling size. This is applied to each secondary home assigned a location zone in the secondary home location sub-model in that year. It determines the number of rooms and quantity of floorspace (in square feet) using the same process used in the primary home dwelling size sub-model as described above.

Primary and secondary home grid cell allocation. This is applied to each home (either primary or secondary) assigning a location zone in that year. It determines the specific grid cell where the home is situated within the location zone it has been assigned. There is an ongoing accounting of the quantity of vacant residential space in each grid cell. The grid cell for the home is selected using a Monte Carlo process, where the selection probability for each cell is proportional to the amount of vacant residential space in the grid cell relative to all the other grid cells of that space type in the zone. The home is then allocated to the grid cell and the quantity of vacant residential space in the cell is reduced by the quantity of residential space assigned to the home in the relevant dwelling size sub-model. Once the quantity of vacant residential space goes negative in a given grid cell, then that grid cell is no longer available in the grid cell allocation process. There is also an ongoing accounting of the total quantity of vacant residential space in each zone, done separately from the accounting of vacant residential space in each grid cell. The primary and secondary home location sub-models described above use these zonal-level total quantities, which ensures that the total quantity of available floorspace in a zone will not go negative even though the quantity of available floorspace in some of its constituent grid cells do go negative. This is done because the grid cells are too small to impose on them the condition that they must contain an integer number of dwelling units. The accounting of integer numbers of dwelling units must be done at a higher level of aggregation – currently at the level of the alpha zones.

Residential space price update. This is applied to each non-zero quantity of each type of residential floorspace in each alpha zone once in each sub-timestep. It makes a one-step adjustment to the unit price in reaction to the vacancy rate relative to a reference vacancy rate value, increasing when the vacancy rate is lower than the reference value and decreasing when the vacancy rate is higher than the reference value, with the provision that the unit price is not allowed to go negative.

The specific size of the one-step adjustment is a function of the values of the vacancy rate and the unit price and a parameter whose value is determined as part of calibration (Hunt et al. 2003; Khan et al. 2002).

Auto ownership. This is applied to each household in the current year, determining the number of private automobiles owned by the household. Specifically, the household is assigned one of the following alternatives regarding the number of automobiles owned: "0", "1", "2" or "3+". A Monte Carlo process is used, with a logit model used to assign the selection probabilities to the alternatives. The utility functions for the alternatives include attributes of the household, such as the number of persons and workers, income, housing type (expressed as "single family unit" or "not single family unit"), level of transit accessibility and previous number of automobiles owned (Portland Metro 2000).

4 Land Development Module

The Land Development (LD) Module provides a representation of the development of space using the system of 30 m x 30 m grid cells covering the model area, microsimulating development transitions occurring in each cell over the period of 1 year. Analogous to the situation with the HA Module, the intention with the LD Module is to perform an endogenous determination of changes in developed space over time and in response to a wide range of potential policy actions involving pricing, regulation and infrastructure in both transportation and land use.

The LD Module considers the full range of possible development types (residential, non-residential), consistent with representation of the full range of activities provided by the Oregon2 Model. This consideration of the full range of possible development types is described here, but with a focus on the residential component consistent with household behaviour being the topic of this chapter.

4.1 Definitions and Categories

Table 2 shows the attributes of individual grid cells tracked in the microsimulation, together with the coding for each. Each grid cell is treated as homogeneous, with only one value for each attribute of the grid cell.

4.2 Process

In a given run of the module, for a particular year, each grid cell is considered in turn starting in the southwest corner and working east and north through the entire model area.

Table 2 Grid cell attributes in LD and HA modules. The 30 m x 30 m grid cells covering the model area are described in terms of these attributes, and the states assigned to these attributes can be changed as part of the consideration of the grid cell in the model

Code	Description		
AZone	Alpha Zone		
DEVTYPE	Development type		
	0 UNDEV		
	1 UNDEVFor (Non-logging forestlands)		
	2 UNDEVAg (Non-productive Ag lands)		
	11 MH (Mobile home residential)		
	12 MF (Multi-family residential)		
	13 AT (Attached residential)		
	14 SFD (Single-family residential)		
	15 RRMH (Rural res. mobile home>=1 acre/unit)		
	16 RRSFD (Rural res. single family >=1 acre/unit)		
	21 FLR accommodation		
	22 FLR depot		
	23 FLR government support		
	24 FLR grade-school		
	25 FLR heavy industry		
	26 FLR hospital		
	27 FLR institutional		
	28 FLR light industry		
	29 FLR office		
	30 FLR retail		
	31 FLR warehouse		
	32 FLR agriculture (productive Ag/range/mining lands)		
	33 FLR logging (logging timberland)		
	41 TP (Oregon2 transportation network ROW)		
	42 Water		
	43 Outside study area		
SQFT	Existing building space (except for Ag/Log lands in land sqft)		
NODE_ID	XXXX	Nearest network node (Fnode)	
DEVELOPED	XX	Summary land classification	
	01	Urban developed	
	02	Urban undeveloped	
	11	Rural developed	
	12	Rural undeveloped	
	13	Rural protected	
DIST2DEV	Distance to nearest urban developed grid cell (1/4 mile ring midpoint, up to maximum 15,000 ft)		
YRBUILT	XXXX	Year of building construction	
SBW	SBW	3-Digit site Prep cost factors where:	
	S= Slope	B= Depth to bedrock	W= Depth to water table
	1. 0–10%	1. 0–18″	1. 0–6″
	2. 10–15%	2. 18–30″	2. 6–18′
	3. 15–25%	3. 30–36″	3. 18–30″
	4. 25–35%	4. Over 36′	4. 30–36″
	5. Over 35%		5. Over 36′
ZONING	100 ureslo	Urban low density residential	
	101 uresmed	Urban medium density residential	
	102 uresmdh	Urban medium high density residential	

(*continued*)

Table 2 (continued)

Code	Description	
	103 ureshi	Urban high density residential
	104 umix	Urban mixed use
	105 umixhi	Urban mixed use high-density (Portland CBD)
	106 ucom	Urban commercial
	107 uindlt	Urban light industrial
	108 uind	Urban any industrial
	109 upub	Urban public (including schools, parks)
	110 uoth	Urban other use
	111 ubigany	Urban big city any use (halo only)
	112 uany	Urban small city any use
	200 rres	Rural residential – low density
	201 rcntr	Rural center – low density mixed use
	202 rcom	Rural commercial
	203 rind	Rural industrial
	204 rpub	Rural public use (including schools, parks)
	205 rreserve	Indian reservation, Military, DOE, COE, BOR
	206 roth	Rural other
	207 rany	Rural unincorporated cities any use
	300 rfor	Forest lands (logging allowed)
	301 ragfor	Agriculture, logging, range, mining lands
	302 rag	Agriculture lands
	303 rrange	Range lands
	304 rmine	Mining lands
	400 xrec	Protected recreation land
	401 xagfor	Protected agriculture, mining and forest land
	402 xcons	Protected natural areas and conservation zones
	403 xhalo	Protected rural land (halo only)
	404 xother	Protected other
	405 xother2	Not resolved (slivers without data, protect)
	500 xwater	Water
	600 xtp	Major transportation ROW
GRIDFEE	Development Fee Scheme	
	XXX	Fee/subsidy scheme

Grid cells with zoning regulations that do not permit any development (a comparatively large proportion of the total) are skipped. When a grid cell is considered, a series of sub-models takes the cell through possible changes in development and updates its attributes accordingly. This series of sub-models is shown in Fig. 7.

The first LD sub-model ("change in development") determines if a change in the development of the cell occurs, including a change from undeveloped to some initial development. Then, in cases where there is to be a change, further sub-models are used to determine what the new development type is to be ("update development type") and how much of it there is to be ("update development quantity"). The age of construction (YRBUILT) for the development of the cell is then updated, and the process moves on to consideration of the next cell.

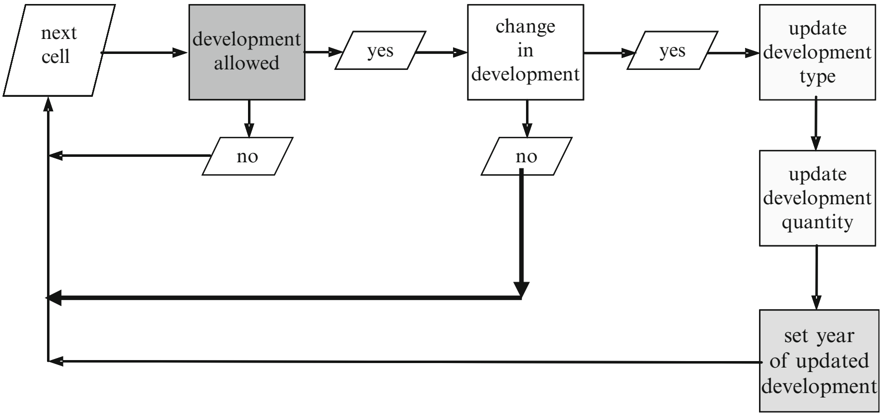
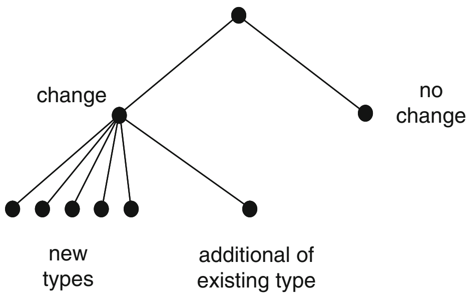


Fig. 7 LD module process diagram. For each year simulated, each grid cell is taken through this process covering aging of existing development and possible changes in development

Fig. 8 Change in development and updated development type choice model nesting structure. The probabilities for changes in development and the resulting updated (new or existing) types are calculated using a two-level nested logit model



The distance to existing development (DIST2DEV) for each grid cell is updated separately after the development simulation has been run for all cells.

The age of construction and distance to development is updated using a rule-based approach. Monte Carlo techniques are used to determine the results for the other sub-models. The selection probabilities for the possible new conditions for the first ("change in development") and the second ("update development type") sub-models are based on a two-level nested logit model with a structure as shown in Fig. 8.

The utility functions in this nested logit model include the net revenue values for the alternative development decisions, calculated by subtracting amortized (re-) development costs from estimated revenues based on current rents identified in the HA (residential) and PI (non-residential) modules. The selection probabilities for the third model, concerning the different quantities of new development ("update development quantity"), are based on a uniform distribution across all quantities from 0 to the maximum allowable developed space accorded under the zoning for the cell.

Each grid cell is considered independently. Interactions between grid cells occur through the rents calculated in the other modules based on the interaction between floorspace supply and floorspace demand.

Considering the LD sub-models in detail:

Change in development. This is applied to each grid cell where the zoning regulations permit change. It determines if the developed space in the cell is to change or remain the same for that year. It uses the first stage (upper level) of the nested logit model with the structure shown in Fig. 8, which includes the log-sum for the full set of allowable development states for the cell as an indication of the expected maximum utility for the "change" alternative. The "lower level" utility functions for these allowable development states include, among other things, the corresponding unit rents as determined in the PI Module for the alpha zone containing the cell, less the corresponding unit construction costs amortized into monthly amounts. The utility function for the "no change" option includes the unit rent for the existing space type in the cell and a term representing the impact of the age of the existing structure to account for the lower appeal (and lower rents and rent-ability) and higher maintenance costs generally associated with older structures. A Monte Carlo process is used to select between the "change" and "no change" options, with the selection probabilities determined using the first stage (upper level) logit model.

Updated development type. This is applied in each case where the grid cell is assigned the "change" option as described immediately above. It determines which of the allowable development states results from the change. It uses the second stage (lower level) of the nested logit model with the structure shown in Fig. 8. The "new types" alternatives all involve demolition of the existing space and construction of new space of a particular type – as allowed by the zoning regulations. The "additional of existing type" alternative represents the option of adding more floorspace of the current type, to more intensely develop the cell without demolishing the existing space – again, as allowed by the regulations. The utility functions for these allowable development options include the corresponding unit rents as determined in the PI Module for the alpha zone containing the cell, less the corresponding unit construction costs amortized into monthly amounts. The unit construction cost for a given development type in a cell includes five terms (a) a site preparation fee based on a weighted scaling of slope, depth to bedrock and water table characteristics; (b) a building construction cost varying by development type; (c) a fee (or subsidy, as a negative fee) specific to the location and (d) a fee (or subsidy) specific to the development type and zoning and (e) servicing costs varying by the development type, the current servicing at the location and the distance to existing services. Also included in these utility functions are both (a) terms that represent the effects of model-area-wide and zonal-level vacancy rates and associated uncertainties regarding future revenue streams for different development states and (b) alternative specific constants for different development states, that can be used to calibrate the rate of development by space type in the entire model overall. The construction costs and rents are calculated based on the maximum allowable intensity of the cell; in more recent developments of the framework in Baltimore (Hunt et al. 2007) the construction costs and rents effects are calculated as an integration over the continuous logit model used for the development quantity decision, which provides for more realistic development patterns if high intensities are allowed but are not realistic due to nonlinear increases in construction costs (or decreases in market rents) with intensity.

Again, a Monte Carlo process is used to select one development type option, with the selection probabilities determined using the first stage (upper level) logit model.

Updated development quantity. This is applied in each case where the grid cell is assigned the "change" option, immediately after the updated development type model has been applied. It determines the quantity (and thus the intensity) of development within a grid cell resulting with the change in development. For new construction, the maximum allowable floor area ratio (FAR) specified in the zoning rules is used as an upper bound on a uniform distribution with the lower bound set at zero. A value for the SQFT grid attribute is then selected from this uniform distribution. For additional construction (when the "additional of existing type" option is selected), the upper bound is the same as for new construction, but the lower bound is the previous value of the SQFT grid attribute for the cell.

In recent developments in a similar model in Baltimore (Hunt et al. 2007), a continuous logit formulation is used, where each infinitesimal intensity option has a utility. This provides additional flexibility and also provides more consistent integral-logsum measures for the development type decision.

Year of construction. This is applied to each grid cell with development occurring in the current year. If there is a change in development and the updated type is new, then the age of construction is reset to the current year. If there is a change in development and the updated type is additional of the existing type, then the age of construction is set to the average of the year of the existing development and the current year weighted by the corresponding quantities.

To ease the computational burdens for the HA and PI Modules, the LD Module updates a database of the developed grid cells – about 25 million out of the 810 million cells covering the entire model area – which allows the HA and PI modules to ignore the vast majority of the grid cells – all those containing the undeveloped land.

5 Parameter Development and Model Calibration

The values for the parameters in the HA and LD Modules are established jointly with the rest of those in the entire Oregon2 model using a three-stage process as illustrated in Fig. 9.

In the first stage, values are developed for certain "S1" parameters. It is unlikely that S1 parameter values will be adjusted as the development and calibration work progresses. In some cases, statistical methods are used to estimate appropriate S1 parameter values; in others, only a single observation is available and direct

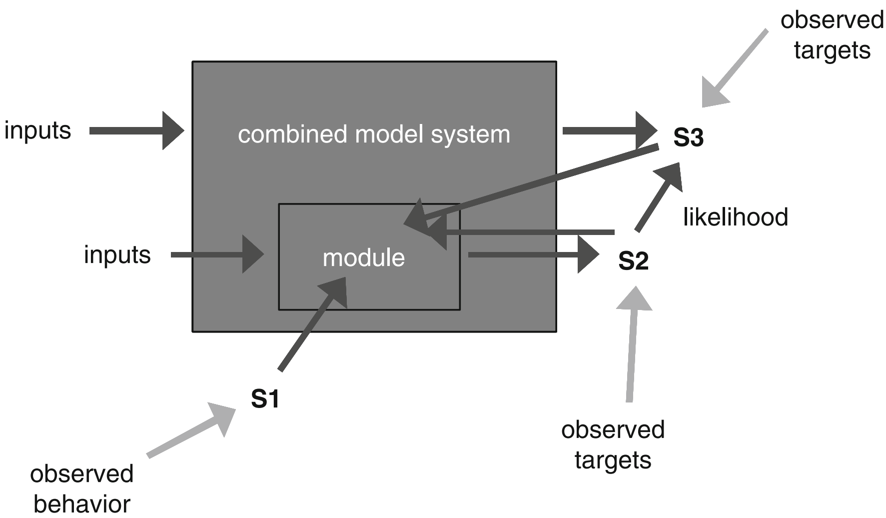


Fig. 9 Approach in development of module parameters. The values for parameters are established in a 3-stage process. In the first stage, the values for the "S1" parameters are determined using external data and statistical methods and remain largely fixed through the rest of the calibration. In the second stage, the values for the "S2" parameters for each module are determined using the fit of the module against targets. In the third stage, the values for the "S3" parameters, which are a selected subset of the "S2" parameters, are determined jointly for all modules using the fit of all modules against targets

methods are used. At this point in the process, it is not necessary that the entire module be run: the components sub-models of the module are being "assembled" and the outputs of the module are not yet being considered.

In the second stage, initial values are established for all the parameters in the module other than the S1 parameters, called the "S2" parameters, considering the fit of the module in isolation. This fit concerns specified targets for outputs from the module, so the module needs to be run in order for it to provide these outputs. Thus, a full set of required inputs for the module needs to be developed, including all those provided by other modules and all those provided exogenously.

In the third stage, the initial values established for certain sets of the S2 parameters are revisited – for all of the modules simultaneously, considering the fit of all modules together, with the full Oregon2 model running, so that inputs to the modules are coming from the other modules in the way they would for a complete model run.

A Bayesian process is used to some extent, to make maximum use of any prior knowledge on parameter values (Abraham 2000). In the estimation of the S3 parameters, the goodness of fit measure guiding the process includes – along with the comparison of module outputs with targets – a comparison of the current values with those determined in the second stage, which amounts to using the results from the second stage as the Bayesian prior distributions for the S3 parameters considered in the third stage (Bard 1974). The initial tradeoffs between matching the various different types of targets and respecting previously established parameter values are guided by (in part notional) estimates of the amount of error in each target and the confidence intervals ideally established during earlier estimations but in some cases based on perceptions. Ultimately, the purpose and use of the model is considered and the tradeoffs between different types of fit and different prior knowledge adjusted accordingly.

Ideally, all parameter values would be revisited at each stage, but this is too open-ended and thus impractical with regard to the resource requirements. An increasing number of parameter values are fixed at each stage, which is from a Bayesian perspective identical to putting an infinite weight on the prior knowledge for that variable.

A weight sensitivity matrix (Abraham 2000) is used to explore the remaining lack-of-fit for the entire model, which can help identify the parameters to focus on in the third stage, which may lead to small changes in the details of the model design and specification.

At the time of writing, the entire system as described is operational and much of the "S1" parameters have been established, but work on the "S2" parameters has yet to begin in earnest.

6 Conclusions

The disaggregate microsimulation treatment of households in the Oregon2 model – as accomplished with the set of seven interconnected modules in the entire system, and with the HA and LD Modules in particular – provides a representation of the changes in demographics and in developed residential space occurring over time and in response to a wide range of potential policy actions involving pricing, regulation and infrastructure in both transportation and land use. At the time of writing, the system is not yet complete in that much of the second and all of the third stages of calibration are still outstanding. But some preliminary conclusions about the design and development of the modelling system, and the representation of households in particular, can still be offered.

The entire framework is clearly an ambitious attempt to incorporate explicit representation of a very extensive range of elements of the entire spatial economic system. Much has been done "from scratch" in the design, the development of software code and in the amassing of the relevant data. And a considerable calibration effort still remains. It still seems to be the case that, overall, the development of the framework is an appropriate undertaking given the very large potential implications of the policy development that can be expected to benefit from the results. The sponsors of the work are to be lauded for recognizing and supporting this long-term view – going well beyond what is typically the case in this regard. But, even with this long-term view, it is increasingly apparent that a staged approach to development is appropriate, providing interim results that can be used before the full system is ready enabling model development to be driven in part, in response to the policy questions and applications of interest. Indeed, pressure to respond to analysis needs within a shorter timeframe led to the development of the Oregon2 Transitional Model, with its more aggregate treatments, discussed above. The Transitional Model has undergone sensitivity testing (Weidner et al. in press) and is beginning to be used for policy analysis.

Perhaps not surprisingly, the run time for the full Oregon2 simulation is an important factor. Some of the modules run very quickly, in a matter of minutes. But the PT module, simulating the activities and trip making behaviour of over five million persons, takes about 3 h to run each year, even after some concerted effort at streamlining the code and parallelization across multiple computers. The HA Module takes a similar amount of time, although it has yet to receive much attention directed at making it run faster. The LD Module requires less time – after the initial work preparing the information for the almost one billion grid cells is done. For a simulation over a multiple number of years, the total run time can become unacceptably large. In order to reduce this total run time, the transport models, including the PT Module, can be run for only certain years, perhaps for 1 year out of every three or five, with the results saved and used as the inputs for the other modules in all intervening years.

The relevant data for the HA and LD Modules have been available, drawing on the US Census, the Panel Survey of Income Dynamics, the Oregon Travel Behavior Survey, Oregon Vital Statistics Reports, and the American Housing Survey. As is to be expected, there have been issues of compatibility for data from different sources collected at different times. Nevertheless, the available data have been enough to produce reasonable behaviour in each of the submodels except for those relating to second homes. In some cases datasets for Oregon were too small or unavailable, as in the use of the American Housing Survey sample for the entire US, and the use of the PSID data. The success with these data sources reflects the fact that, except in the case of secondary homes, the nature of the available data was taken into account in the design of the model.

The system of grid cells has given rise to practical challenges because of its very large number of cells. Yet this level of resolution is required if the system is to support some of the policy analysis as intended. Much of the computational burden is reduced by completely ignoring the 68% of grid cells that do not allow any future development (large regions of undeveloped areas, such as in the Cascade Mountains), and by blocking together (and considering in groups) the 19% of grid cells that only allow land-intensive agricultural and forestry development. The initial work to "fill" all of the grid cells with the pertinent information for the starting year, and to deal with the inconsistencies and discrepancies in the components of information from different sources at this level of resolution, was so substantial that it was a large part of the reason efforts were diverted to develop an interim version of the model without the grid cells and with more aggregate treatments of land development and household demographics (Abraham et al. 2005).

Some submodels are still overly simple. For instance, the intensity of development decision in the model is a simple uniform allocation, while in the real world this decision rests on the willingness-to-pay for lower intensity floorspace on the demand side (related to such things as the amenity value of associated undeveloped land, the distance to neighbors, and a higher value of ground floorspace) and the additional construction costs associated with higher intensity on the supply side. It is thought that the willingness-to-pay for lower intensity should, eventually, come from future enhancements in the floorspace demand models in HA and PI. It may turn out, however, that more realistic intensity of development decisions can be simulated through a modified supply side representation in LD that would be much simpler, but not as behavioural, as a demand side representation in HA and PI.

From a transport modelling perspective, the primary purpose of the HA and LD Modules is to forecast the changes in the distribution and location of population. They are influenced by transportation system attributes both directly and indirectly. Direct influences include the travel conditions for the implied movement between homes, workplaces and second homes in the utility functions for primary home choice, home-anchored job search, and secondary home choice. Indirect influences include the relationships between floorspace demands, supplies and prices across space and the differences in these demands arising because of the relative accessibilities and the varying sensitivities of different activities to these accessibilities – ultimately incorporating the establishment of the full activity system, including the full range of job opportunities and wage rates, as represented in the rest of the modelling system.

In a dynamic time-series simulation like the Oregon2 model, every attribute value used as an input in each time step needs to be considered and possibly updated in every time step. That is, more specifically, all the person and household attributes used in the models of travel and transport (in the PT and TS Modules) need to be considered and possibly updated in each time step – which is part of the reason why the HA and LD Modules address so many attributes in their various sub-models; requiring a substantial amount of demographic transition modelling that, in and of itself, is not directly related to any of the transportation-related policy issues that might be considered with the model. An alternative approach avoiding much of this demographic modelling, in cases where the intent is to develop a modelling system for addressing transportation policy only, would be to keep household attributes other than location fixed over time and only forecast location changes for jobs and housing. This would be much simpler than the approach adopted here. It would suffer by missing a primary reason that local (neighbourhood) demographics change – the aging and changing of lifestyle of existing residents.

The HA and LD Modules interact with each other over time primarily through the markets for the different types of residential floorspace. The development of these two Modules in parallel as process-oriented microsimulations forced consideration of the temporal dynamics of the joint system represented by the two Modules without resorting to the use of a full short-run equilibrium framework. The resulting approach – where supply and demand are linked via prices, with these prices updated in the direction towards but not all the way to equilibrium based on excess demand – has be found to provide realistic looking patterns in aggregate over the long term in trials of a small test system. This approach might be extended in the future to other longer term supply–demand interactions, including the labor markets. If the 1-year time step is abandoned (in favor of, say, a discrete event simulation approach with all modules operating simultaneously) the approach could also be applied to shorter-term supply–demand interactions, such as the demand for link travel.

With the explicit treatment of household and land development processes at the individual grid cell level in the HA and LD Modules, it is possible to draw on knowledge of these processes available from other research work in ways not possible with more aggregate (and thus more synthetic) representations: household location and auto ownership decisions can be tied to individual, mode-specific accessibilities; land development decisions can be influenced by local prices and area-wide vacancy rates. Nevertheless, at many points during the design and development of these modules it became clear that knowledge of the relevant decision-making processes is very limited.

Throughout the entire modelling system there is extensive use of Monte Carlo processes with logit choice models. This is not based on any strong conviction among the developers that in the real world decision-makers use the process of full-information, compensatory trade-off assessment of discrete alternatives that underlies the logit formulation. In fact, the expectation among the developers is that the actual decision-making behavioural mechanisms are simple, rule-based search processes with severely limited information. Yet lack of knowledge about the specifics of the actual behavioural mechanisms, about the nature of any search processes that might be involved, led the developers to "fall back" on that with which they were more familiar – the logit model and the use of the probabilities it provides within a Monte Carlo approach. In the longer term, as more understanding of the actual behavioural mechanisms is developed, to the extent that these are based on simple rules and limited knowledge, a greater use of explicit representation of these mechanisms can be expected not only to provide a greater fidelity in representation but also to reduce the amount of information processing and thus to reduce the time it takes for the modules to run.

At this point, the next steps to be done are to calibrate the modelling system using the process described above, and to apply the resulting model in the analysis of policy in Oregon. Then it will be possible to gain more complete and valid insights into the extent that the entire system, including its representation of household behaviour, helps provide better guidance in policy development. Many different enhancements are possible, but it is more appropriate to pursue the completion of the system first, in order to demonstrate and test its functionality and ability to analyze policy.

Acknowledgements The work described here was sponsored by the Oregon Department of Transportation. The development of the HA and LD Modules, along with the other modules in the Oregon2 model system, has been done by a series of teams led by PB Consult, with Hunt Analytics Incorporated, HBA Specto Incorporated, EcoNorthwest and the University of Washington playing key roles as sub-contractors. More detailed descriptions of this work and resulting system are provided in project reports (Hunt et al. 2001; PBConsult et al. 2003a, b).

References

Abraham JE (2000) Parameter estimation in urban models: theory and application to a land use transport interaction model of the Sacramento, California Region. Dissertation, University of Calgary

Abraham JE, Weidner TJ, Gliebe J, Willison C, Hunt JD (2005) Three methods for synthesizing base-year built form for use in integrated land use – transport models. Transp Res Rec 1902:114–123

Bard Y (1974) Nonlinear parameter estimation. Academic Press, New York

Beckman RJ, Baggerly KA, McKay MD (1996) Creating synthetic baseline populations. Transp Res 30A:415–429

Hunt JD, Donnelly R, Abraham JE, Batten C, Freedman J, Hicks J, Costinett PJ, Upton WJ (2001) Design of a statewide land use transport interaction model for Oregon. In: Proceedings of the 9th World conference for transport research. Seoul, South Korea, July 2001

Hunt JD, Khan J, Abraham JE (2003) Micro-simulating firm spatial behaviour. In: Proceedings of the 8th international conference on computers in urban planning and urban management. Sendai, Japan, May 2003

Hunt JD, Abraham JE, De Silva D, Zhong M, Bridges J, Mysko J (2007) Microsimulating space development in Baltimore. In: Proceedings of the 12th international conference of the hong kong society for transportation studies: transportation systems engineering and management. Hong Kong SAR, China, December 2007, pp 513–522

Khan JAS, Abraham JE, Hunt JD (2002) Micro-Simulation of the activities and locations of business establishments. Paper presented at the 49th Annual North American Meetings of the Regional Science Association International, San Juan, Puerto Rico, 14–16 November 2002

Metro P (2000) Transport demand model auto ownership model description. Portland Metro Transportation Department, Portland OR, USA

National Center for Health Statistics (1999) National vital statistics reports, vol 49(1) Data on number of babies born in Oregon. United States Bureau of Statistics, Washington DC, USA

National Center for Health Statistics (2000) National vital statistics reports, vol 49(1) Data on frequency of male births in United States. United States Bureau of Statistics, Washington DC, USA

Oregon Department of Human Services (1995) Oregon vital statistics report, vol 1, Give birth probability data. Oregon Department of Human Services, Salem OR, USA. http://www.dhs.state.or.us/publichealth/chs/vol1.cfm. Accessed 2001

Oregon Department of Transportation (1994) Oregon travel behavior survey. Oregon Department of Transportation, Transportation Planning Analysis Unit (TPAU), Salem OR, USA. Data for Municipal Planning Organizations (MPOs), Portland Metro, Salem/MWVCOG, Eugene/LCOG, Medford/RVCOG. http://www.oregon.gov/ODOT/TD/TP/TMR.shtml#96_OR_Travel_Behavior_Surveys. Accessed 2001

Oregon Department of Transportation (1996) Oregon travel behavior survey. Oregon Department of Transportation, Transportation Planning Analysis Unit (TPAU), Salem OR, USA. Data for 8 Rural Counties: Clatsap, Coos, Deschutes, Josephine, Klamath, Lincoln, Malheur, and Umatilla. http://www.oregon.gov/ODOT/TD/TP/TMR.shtml#96_OR_Travel_Behavior_Surveys. Accessed 2001

Oregon Department of Human Services (1998) Oregon vital statistics report, vol 1, Death probability data. Oregon Department of Human Services, Salem OR, USA. http://www.dhs.state.or.us/publichealth/chs/vol1.cfm. Accessed 2001

PBConsult, HBA Specto, EcoNorthwest (2003a) Oregon2 model development, HA module description at finalization. First Draft Report. Submitted to Oregon Department of Transportation, Salem OR, USA

PBConsult, HBA Specto, EcoNorthwest (2003b) Oregon2 model development, LD module description at finalization. First Draft Report. Submitted to Oregon Department of Transportation, Salem OR, USA

University of Michigan (1981–1993) Panel survey of income dynamics (PSID). Inter-University Consortium for Political and Social Research data archive. University of Michigan Survey Research Center, Ann Arbor MI, USA. http://simba.isr.umich.edu/. Accessed 2001

US Census (1990a) Census of population; Public use microdata samples (PUMS) dataset. United States Government Printing Office, Washington DC, USA http://www.census.gov. Accessed 2001

US Census (1990b) Census of Population; Summary Tape File (STF3) dataset, Series CD90-3. United States Government Printing Office, Washington DC, USA. http://www.census.gov. Accessed 2001

US Census (2002) Current housing reports, Series H150/01, American housing survey: 2001. United States Government Printing Office, Washington, DC, USA

Weidner TJ, Knudson RA, Picado R, Hunt JD (2010) Sensitivity testing with the Oregon Statewide Integrated Model. Transportation Research Record 2133:109–122

Williams HCWL (1977) On the formation of travel demand models and economic evaluation measures of user benefit. Environ Plann 9A:285–344

The Residential Choice Module in the Albatross and Ramblas Model Systems

Theo Arentze, Harry Timmermans and Jan Veldhuisen

Abstract The focus of this chapter is on the residential choice component in the Albatross and Ramblas model systems. Both models are primarily activity-based models of transport demand. Their prime goal is to predict activity–travel patterns and associated traffic flows. The distribution of residential land use, in terms of households and persons, is exogenously given.

Most progress to date in terms of actual software development has been completed in the context of Ramblas. It contains a module for modelling residential choice behaviour that is used to predict the choice of residential zone for people moving house and newcomers in the housing market. Simultaneously, the properties of the dwelling stock are updated. Residential preferences measured in theNational Housing Survey are matched against vacant dwellings in the market. These preferences are measured using a compositional stated preference approach, but alternatively any conjoint preference approach could be used in principle.

1 Introduction

The aim of this chapter is to summarize research activities of the Urban Planning Group of the Eindhoven University of Technology related to the modelling of residential choice behaviour. These research activities have been developed along two separate lines of research. First and most importantly, residential choice has been an important domain of application to elaborate conjoint preference and choice models that confusingly have been called stated preference models in the transportation literature. Starting in the early 1980s (Veldhuisen and Timmermans 1984), the reliability and validity of conjoint preference models was tested and this interest evolved into a continuous stream of research activities that extended and

T. Arentze (✉), H. Timmermans, and J. Veldhuisen
Eindhoven University of Technology, Urban Planning Group, Eindhoven, The Netherlands
e-mail: t.a.arentze@bwk.tue.nl

F. Pagliara et al. (eds.), *Residential Location Choice*, Advances in Spatial Science,
DOI 10.1007/978-3-642-12788-5_10,

improved this modelling approach. Originally, one of the key problems in developing conjoint preference models of residential choice was how to include the many attributes that influence the residential choice decision. The hierarchical information approach, originally developed for preference tasks (Louviere 1984), was generalized to hierarchical conjoint *choice* models (Timmermans 1989; Louviere and Timmermans 1990). Later, an improved method, named integrated choice experiments, was introduced in the literature (Oppewal et al. 1994), and this method was tested to model residential choice behaviour (Van de Vijvere et al. 1997). Other extensions concerned the development of a *context-dependent* conjoint choice model of residential choice behaviour (Timmermans and van Noortwijk 1995), and a model incorporating the similarity between attributes (Timmermans et al. 1996). Finally, the modelling approach was generalised to the modelling of group preferences and group/family choice as opposed to individual choice behaviour, and it has been argued that residential choice behaviour might be better viewed as a group/family decision than an individual decision (Timmermans et al. 1992; Molin et al. 1999).

In addition to such research focusing on a particular type of behaviour, such as residential choice, the Urban Planning Group has been active in developing integrated models. Examples of such models are the activity-based models Albatross and Ramblas. The latter model can best be viewed as a modern version and extension of "A Regional Location Model", which was developed in the mid 1970s (Veldhuisen and Kapoen 1977). This model allocates different land use across space and was similar in scope and objectives as the conventional land use – transport models. It differed in that the allocation of residential land use was not based on observed behaviour but rather on people's measured preferences. In the late 1990s, the model was revitalised and given the new acronym Ramblas. As will be discussed later in more detail, the transport component was new, being based on observed activity–travel patterns. In addition, micro-simulation using readily available empirical data was employed. It is a data-driven model, there is no attempt to interpret and generalise the data in terms of some underlying theory.

In contrast, Albatross is theory-driven. The model was originally developed for the Dutch Ministry of Transport and is best viewed as a rule-based, computational process model of transport demand. A series of rules is used to predict which activities will be conducted where, when, for how long, with whom, and the transport mode involved. Residential land use is an exogenous variable in the model. However, as part of the Amadeus research programme (Timmermans et al. 2002), partly completed, partly ongoing and planned work on Albatross includes spatial population forecasts as a function of among other housing plans, and forecasts of land uses, including residential land use. The two activity-based models, which will remain different in the way they model activity–travel patterns, will likely share much of the way in which residential choice is treated. Therefore, the two models are discussed simultaneously in the present chapter.

The chapter is organised as follows. First, we will discuss in more detail the motivation, underlying principles, scope and structure of Ramblas, followed by a similar discussion of the Albatross model system. Next, we will discuss the way in which components that were developed to model residential choice can be incorporated in the models. This discussion involves how a baseline synthetic population is derived from data on the distribution of dwellings in the study area and how housing choice of movers and newcomers is modelled. Finally, we will draw some conclusions and briefly discuss plans for future research.

2 Ramblas

To give an appropriate framework for the discussion of the residential choice module, we first briefly summarize the core of both models: the simulation of activity–travel patterns. Ramblas has been developed to explore the possibilities of developing a micro-simulation model that uses nationally available, easy accessible, official statistics only, and that is based on simple and easy to comprehend principles as opposed to specific modelling techniques. The model has been developed primarily to estimate the intended and unintended consequences of planning decisions related to land use, building programs and road construction. Its main purpose is to predict the spatial distribution of individuals' activities and related traffic flows, given a forecasted spatial distribution of dwellings, households, firms, and the transport network.

Assume that a list of individuals with a set of characteristics is given. Later, we will discuss in more detail how this is done. For each individual, an activity agenda is created for some specific day of the week by matching the socio-demographic profile of each individual to nationally available data on time use. Individuals are classified according to 26 segments, based on gender, age, employment status and education, and five types of municipalities. An activity agenda is created by identifying the relevant segment and drawing at random an activity pattern from the national database of that segment. Seven types of activities are distinguished: work, child care, shopping, personal/medical care, school or study, social participation and social contacts.

For each out-of-home activity, transportation mode choice is simulated by drawing at random from the corresponding conditional probability distributions, created from the national time use survey. Once activity agendas and transportation modes are known, the next step in the micro-simulation addresses the problem of how this agenda is executed in space. In the case of the work activity, it is assumed that the travel time observed in the diary constitutes the time people are willing to travel to work, given the transportation mode involved. Destination choice for the work activity is simulated by drawing at random without replacement a job location from the total number of available jobs in the region, which is delimited by this maximum travel time. In the case of school, it is assumed that children going to elementary schools invariably choose the school nearest to their home. For students going to secondary schools, it is assumed that their action space is defined by an area of 45 min of bicycling time. Schools are drawn at random from this action space. The same principle is used for students going to schools of higher education, but in this case the distribution of employment in higher education is used as the distribution from which the school is sampled. The latter principle is also used to simulate destination choice for shopping and services. The destination is drawn at random from the distribution of employment in the relevant sectors. As for social activities, the distribution of the population rather than the distribution of employment is used as the distribution from which the destination is sampled.

The above simulation process results in an origin and destination, plus a simulated transportation mode, for each trip. If the transport mode involves the car, route choice behaviour is also simulated, assuming that individuals take the shortest route, in terms of travel time. These trips are loaded onto the network. The "speed-flow" method is used to calculate the required travel time. Given the arrival time at the destination, the departure time is then calculated. The simulation process thus results in an estimate of traffic flows on the network for every moment of the day.

3 Albatross

The purpose of Albatross is to predict which activities will be conducted where, when, for how long, with whom, and the transport mode involved. Thus, its objectives are quite similar to those underlying Ramblas. It does however involve more choice aspects, and more (personal, household, spatio-temporal and institutional) constraints. Route choice is not an integral part of the model yet, but should be handled by another model, but an innovative approach is currently under development (Arentze and Timmermans 2003). The model system has been developed for the Dutch Ministry of Transport, Public Works and Water Management in the context of a research and development programme that aims at exploring new ideas and methodologies for transport planning.

Although the two models share this common purpose, they represent extreme examples of activity-based models at the opposites of the spectrum. Ramblas is primarily data-driven. Distributions and conditional probabilities observed in national data sources are used to simulate activity–travel behaviour at the local level, at best correcting for known local data. This is no attempt to capture any structure in this data in terms of an underlying theory or algebraic or rule-based model. In contrast, Albatross is theory-rich, and represents an attempt to extract context-dependent choice rules from activity–travel diary data, collected specifically for developing the model, to simulate activity–travel patterns.

The core of the model is a scheduling agent, which generates a schedule for each individual and each day and consists of two components. The first component generates an activity skeleton consisting of fixed activities and their exact start time and duration. Given the skeleton, the second component then determines the part of the schedule related to flexible activities to be conducted that day, their travel party, duration, time-of-day and travel characteristics. Both components use the same location model component determining the location of activities. All three components assume a sequential decision process in which key choices are made and pre-defined rules delineate choice sets and implement made choices in the current schedule.

The skeleton model determines activity patterns on a continuous time scale. It consists of several sub-processes including: determining the pattern of sleep activities, determining the pattern of the primary work/school activity, determining the pattern of secondary, fixed activities and determining the location of each fixed activity episode. The model chooses the end time of the morning sleep episode and the start time of the evening sleep episode. The primary work/school activity has maximally two episodes and a minimum duration of 1 h per episode. The pattern is defined by decisions about the number of episodes, start time, duration(s) and inter-episode time. Work/school activities with shorter duration are treated as a separate category of secondary fixed activities in the next step.

The location component chooses locations in descending order of priority of fixed activities. For each activity, the choice set consists of all 4PCA's (four position postal code area, if which approximately 4,000 exist) in the Netherlands. First, the model chooses the municipality and next a 4PCA within the chosen municipality. At both levels, the model determines a choice by increasingly narrowing down the choice set in a number of steps. For the choice of municipality at the highest level, the first decision determines whether the activity (episode) is conducted within or outside the home municipality of the individual. If the last option is chosen, the choice of a municipality follows from a choice of an *order* and distance band. Five orders are distinguished based on population size. Given the order, the choice of a distance band follows. The combination of order and distance band tends to reduce choice sets strongly. If there are still multiple alternatives left, the model selects a municipality semi-randomly. For the choice of a zone (i.e., a 4PCA) within the chosen municipality a similar logic is used.

All other choice facets are also modelled using a decision table formalism for choice rules. This set of decision tables is partly linked in the sense that the outcomes of one or more previous decision tables in the assumed scheduling model are input to subsequent tables. The complete Albatross system consists of 1,687 choice rules to simulate activity–travel patterns. Full details of the model are provided in Arentze and Timmermans (2000, 2003).

4 Treatment of Residential Choice Behaviour

The spatial distribution of residential land use plays a double role in the simulation of activity–travel patterns in both models. First, both models assume the construction of a synthetic baseline population at the start of the simulation period. To that effect, the number of individuals and their values on a set of sociodemographics in each postal area are predicted, reflecting the spatial distribution of residential land use. This distribution influences the activity agendas and the spatial–temporal constraints underlying the models. This data can be exogenous to the models, implying that the relevant distributions should be based on an external model or data source and the creation of the synthetic population takes places at each simulation run.

Secondly, residential land use is an integral part of the dynamics in the model systems. In this case, the aging and redistribution of the population, partly reflecting residential choice behaviour, is internal to the model system. In this case, a special sub-model or module predicts housing choice behaviour as a function of socio-demographics, characteristics of the available dwelling stock, characteristics of the transport network, and possibly activity agendas.

4.1 Creation of a Synthetic Baseline Population at $t = 0$ Reflecting the Spatial Distribution of Residences

A synthetic population is represented in both systems in terms of a multiway attribute frequency table. Known demographics for the study area, based on official statistics, define the marginals of the table and the sample the initial cell proportions. The models determine cell proportions that are consistent with both sources of data. Every cell in the table represents a unique combination of attribute levels. If N is the number of cells of the multiway table for zone i and W_{ij} the number of households in cell j, the system generates N households with multiplication factors W_{ij}. Thus, the population of zone i is represented by a N-vector $\boldsymbol{W}_i$, $\forall i$.

The set of attributes and attribute levels that describe the synthetic population are those used in the model to simulate activity–travel pattern. In addition, household attributes are chosen such that individuals can be derived. The two models differ in terms of the specific methodology and data sources that are used to create this baseline population.

Monte Carlo simulation is used in Ramblas. It starts with the population matrix according to age (in years), gender and marital status available for each municipality of the Netherlands, using data published by the Central Bureau of Statistics. This matrix includes the vector of married women. Using the National Housing Survey, household characteristics are added to each married woman by drawing at random from the set of households that have a woman of the same age. This procedure results in number of married men and a set of children by age and gender. These simulated numbers differ from the given population matrix and are therefore corrected to fit the observed number of married men according to age and the children according to age and gender. The correction is based on known distributions of married men by the age of the spouse and those of children by the age of the mother. The surplus of men in every class, respectively children, is reallocated at random to other age cohorts and gender (in the case of children). The case of other household types is straightforward.

The creation of synthetic populations in Albatross differs in a number of regards. First, in addition to age, gender and marital status, household type (single non-worker, single worker, double non-workers, double one-worker, double two workers), socio-economic class (very low household income, low, average, high), number of cars in household, availability of car for person (is capable of using car), and work status of person are used. Secondly, these socio-demographic profiles are not explicitly linked to housing characteristics. Thirdly, whereas Ramblas is based on the National Time Use Survey and The National Housing Survey, Albatross used the National Travel Survey and the Population Data of the Ministry of Transport. These differences reflect the idiosyncracies of the models and the principal's need to use a common data set for different projects.

In addition, Albatross uses a more formal approach to create the synthetic baseline populations. Instead of Monte Carlo simulation, iteratively proportional fitting is used to create multi-way tables. IPF assumes an $I_1 \times I_2 \times ... \times I_m$ table with initial cell counts $m_{i_1...i_m}$and marginal counts $C_{1(i_1)}, C_{2(i_2)}, ..., C_{m(i_m)}$ as given, where I_j represents the number of levels of the j-th attribute, i_j is the i-th level of the j-th attribute, $m_{i_1...i_m}$ is the count in cell $i_1 \times i_2 \times ... \times i_m$ and$C_{j(i_j)}$ is the total count of the i_j-th level of the j-th attribute in the target population. Adjustment of a cell count $m_{i_1...i_m}$given marginal count $C_{j(i_j)}$ is according :

$$m'_{i_1....i_m} = m_{i_1....i_m} \frac{C_{j(i_j)}}{\sum_{i_1...i_{j-1},i_{j+1}...i_m} m_{i_1....i_m}}. \tag{1}$$

This operation is repeated for every margin and every cell until convergence is reached. Although the actual method differs, the two approaches will yield the same results, within some margin, as long as the Monte Carlo simulation is based on the proportionality assumption.

The difference in choice of method reflects the specific purpose of the model. The main advantage of the IPF-method is that it is easy to derive the multi-way table that is consistent with some correlation structure, allowing the creation of a synthetic population that is consistent with an assumed future population according to some scenario.

Households need to be allocated to the existing dwelling stock. In both models, households are spatially allocated to the existing dwelling stock given the following constraints (1) for each zone i the number of dwellings equals the number of households and (2) the allocation is consistent with a dwelling-type $\times$ household-type matrix. The dwelling stock in zone i is represented by a Q-vector $\boldsymbol{V}_i$ representing the distribution across Q dwelling types. The allocation step results in a new $N \times Q$ dwelling occupancy matrix, say $\boldsymbol{Z}_\mathrm{i}$.

Again, the two models differ in terms of the methodology used to this effect. Ramblas again uses Monte Carlo simulation. The residential preference of a household for type of dwelling is drawn at random from the National Housing Survey, given the household sociodemographics and these preferences are matched to housing characteristics. Households are allocated at random to dwellings that qualify. Households that are not allocated (mainly unmarried young adults) are assumed to share a dwelling. In contrast, for determining $\boldsymbol{Z}_i$, Albatross uses an

Iterative Proportional Fitting method, whereby initial cell proportions are based on observed residential preferences and marginals are given by $\boldsymbol{V}_i$. Consequently, the resulting matrix is consistent with preferences at the dwelling type level and distribution $\boldsymbol{V}_i$. Note that accessibility is not a variable, influencing the allocation of household across space. One of the reasons for this is that accessibility has consistently found to be a relatively unimportant factor in the residential choice decision (Molin and Timmermans 2003).

4.2 Dynamic Residential Choice Behaviour

When the models are used as explained above, they are applied in a static fashion. Exogenous data is used to prepare the synthetic population that in turn constitutes the necessary input to the simulation of activity–travel patterns. This approach would suffice if the aim of the model application is to predict the cross-sectional implication of land use or transport policy on any given point in time. If, however, the goals would be to trace the policy effects over time, then either the above procedures should be repeated for the sequence of year in the forecasting period, using exogenous data, or an internal accounting and residential choice module is required. Ramblas has been fully developed in this regard, the implementation in Albatross is in progress.

Such a module requires both an accounting system, simulating the transitions between household types, and data, reflecting planning measures related to the construction and demolition of dwellings. Users of the models can provide the latter information, but the New Map Foundation also collects such data for The Netherlands. Thus, let there be given $\boldsymbol{V}_{i,t}\ \forall\ t$ based on a given scenario of housing development programs to be realized at the end of t. A housing development program specifies new construction and demolition and is specified for every zone i in terms of a $d\boldsymbol{V}_{i,t}$ such that $\boldsymbol{V}_{i,t+1} = \boldsymbol{V}_{i,t} + d\boldsymbol{V}_{i,t}\ \forall\ t$.

Transitions between household types can be conceptualised (the actual algorithms are based on micro-simulation and agent technology) in terms of a transition matrix. This matrix is a $(N + 1) \times (N + 1)$ matrix, whereby N is the number of cells in multiway table $\boldsymbol{W}_j$. The extra row represents the new households of type j, the extra column the dissolution of existing households of type j and the remaining $N \times N$ cells the transition of household type j to household type j'. If $\boldsymbol{M}$ represents the $N \times N$ transition matrix, the distribution of household types at $t + 1$ may be found by: $\boldsymbol{W}_{j,t+1} = \boldsymbol{W}_{\boldsymbol{j},\boldsymbol{t}} \times \boldsymbol{M}$.

To simulate residential mobility for the simulated individual households, the National Housing Survey is used to determine the number of households searching for a new residence at $t + 1$ as is denoted as $w\boldsymbol{V}_{i,t+1}$. This number includes new households and existing households, wishing to change residential zone or dwelling type. For each zone, the stock of vacant dwellings is determined. The vacant stock is defined as $\boldsymbol{A}_{i,t+1} = \boldsymbol{L}_{i,t} + s\boldsymbol{V}_{i,t} + w\boldsymbol{V}_{i,t+1}$ (all terms are defined as Q-vectors), where $\boldsymbol{L}$ is a surplus, $s\boldsymbol{V}$ the mutation of the dwelling stock during time period

$[t, t + 1]$ and $w\boldsymbol{V}$ are households searching for a new dwelling. Note that the last term preliminary "removes" people that are searching for a new dwelling from their current homes.

The total demand for dwellings, $w\boldsymbol{V}_{i,t+1}$, and the total supply of dwellings, $\boldsymbol{A}_{i,t+1}$, is known as a result of the previous steps. In the final step, the model simulates the allocation of households in array $w\boldsymbol{V}_{i,t+1}$ to the vacant dwelling. Those in array $w\boldsymbol{V}_{i,t+1}$ who are not successfully allocated "return" to their current dwelling. The result of this step is a new dwelling occupancy pattern $\boldsymbol{Z}_{i,t+1}$ and non-occupancy pattern $\boldsymbol{L}_{i,t+1}$.

Ramblas uses data on residential preferences of the various household types from the National Housing Survey and Monte Carlo simulation techniques to reallocate households. A multinomial logit model is used to predict the probability that a moving household prefers a housing type as a function current housing type, type of municipality and size of the household. Note again that transport considerations do not play a role in relocating households.

4.3 Reflection and Future Work

It may be relevant to put this discussion in a broader framework. As discussed in detail in Timmermans et al. (1994) a variety of self-explicated compositional, decompositional and hybrid stated preference methods are available to measure residential preferences. Unless one arguably dramatically restricts the number of influential attributes, it is impossible to estimate individual-level utility functions. Consequently, the very popular multinomial logit (MNL) model is typically estimated at the segment or aggregate level. Now, if an aggregate model is used to simulate individual behaviour, implicitly or explicitly it is assumed that the sample is homogeneous. This would mean that although we have developed a summarising function to represent the data, which looks impressive, by definition the predictive error will, ceteris paribus, be larger as no account was taken of the heterogeneity in residential preference. If the range in the estimated part-worth utilities of a particular attribute in a conjoint experiment is equal to zero, it either means that the attribute is not important or that the preferences of sub-samples counterbalance (or some combination of these). One way of incorporating heterogeneity in the simulation would be to sample from the error distribution, but of course this would potentially bring the simulation very close to the original data.

There is also the issue of validity. If we accept that the MNL is a reasonable model to estimate residential preference/utility functions using experimental design data (although as discussed in the introduction we have developed more advanced, less rigorous models), this does not necessarily mean that the MNL is also a reasonable model to simulate the residential decision making and choice process beyond the experimental task in the real world. In fact, it might be argued there is a significant discrepancy between the experimental task of choosing between two attribute profiles and choosing a house in the real world. Individuals often have limited and imperfect knowledge about choice alternatives in real world markets, their choice is risky as others might buy the candidate house, the choice set changes on a minute-by-minute basis, and the housing search process involves time, effort and cost, implying that individuals and households may act sub-optimally and accept a dwelling that does not maximise their utility. If this argument is accepted, the MNL may be too simple and a more sophisticated model may be required.

The composition of choice sets constitutes another operational and theoretical problem. The predictions of the MNL model will depend on the size and composition of the choice set. The choice set may consist of thousands of alternatives, creating not only operational problems, but choice sets of that size are unrealistic. In reality, movers likely consider only a few options. Moreover, the IIA property underlying the MNL model will not be satisfied, questioning the validity of the approach to estimate MNL models in conjoint choice experiments and applying these models to predict residential choice behaviour and articulating the need to develop and explore alternative modelling approaches, that better mimic the actual decision making process.

Given these considerations and the computational process nature of Albatross, the following module is currently considered to be implemented. A distinction is made between the decision to become engaged in search for a new dwelling, spatial search and choice. As for becoming active, we assume that people may become active when they wish to start a new household, are dissatisfied with the current dwelling, change work location, enter a new life cycle and mimic moving peers in their social network. The last factor is included to account for a mechanism whereby a wish to move is triggered by moves of peers of the household under concern. These events may be related. That is to say, changing work location, entering a new life cycle or social mechanisms may induce dissatisfaction with the current dwelling. We define dissatisfaction as a household's expectation that a change of residence can improve residential utility. Therefore, degree of dissatisfaction is not only a function of attributes of the current dwelling, but also of a household's assessment of the current housing market within the relevant segment. Thus, the criterion here is not disutility, but marginal utility (possible improvement). A positive marginal utility is not a sufficient condition for triggering search. Inertia and the assessed effort involved in searching and movement create a threshold. Moreover, housing markets are not static, but change over time. If the disutility of staying is compensated by the increase of expected utility of entering the market at some later moment in time, a rational household will remain passive until that moment.

Given these considerations, the condition triggering search in its most generic form can be written as:

$$\max T\left(U_T^{move} - \int_t^T U^{stay}(t)dt \right) > c, \tag{2}$$

where T is the time moment of becoming awake, U_T^{move} is the *expected* utility of moving at time T, t is (continuous) time, $U^{stay}(t)$ is marginal *dis*utility of staying and c is a threshold determined by inertia, uncertainty and effort involved in searching and moving. The integral assumes continuous time. In a discrete-time formulation, the integral symbol is replaced by a sigma. For finding appropriate equations for each of these terms, economic theory (of investment decisions) is relevant in as far as bounded rationality is taken into account. It is also worth noting that dissatisfaction as conceptualised here also covers cases where a willingness to move is activated solely by existing supply (i.e., by seeing a house of one's dream by accident). Such an event would raise U^{move} in (2) and increase the probability of moving.

Once an individual or household has become active, a process of (spatial) search is triggered. Like any behavioural model of choice-set formation, a two-staged process is assumed:

$$\Pr(i \in \mathbf{I}) = \Pr(i \in \mathbf{K})\,\Pr(i \in \Re), \tag{3}$$

where i is an index of houses on offer, $\boldsymbol{I}$ is the choice-set of the (awakened) household under concern, $\boldsymbol{K}$ is the known set of houses on offer and $\Re$ is the consideration set generically defined by means of a set of elimination rules. The first term on the RHS of (3) is a function of a set of factors determining the probability that an offer reaches the household passively (e.g., through spatial interaction, social interaction, media, professional advisers) or actively (search in a strict sense). It is assumed that:

$$\Pr(i \in \mathbf{K}) = f(\mathbf{A}, \Im, \mathbf{S}), \tag{4}$$

where $\boldsymbol{A}$ is the current action space of the household (defined as a set of locations), $\Im$ the social network in which the household takes part (defined as a bi-directional graph) and $\boldsymbol{S}$ the search space (defined as a set of locations). Hence, the first two terms correspond to a passive mode and the last term to an active mode of search. Action space comprises all the activity locations and travel routes of (individuals within) the household and, therefore, is known by the system. $\boldsymbol{S}$ can be defined by means of a set of screening rules selecting locations that meet some preferred characteristics that follow from the choice model. Obviously, $\boldsymbol{A}$ and $\boldsymbol{S}$ may overlap and the overlapping subset will have an increased probability. On the other hand, social network $\Im$ should take into account differential probability of exchanging information between actors in the system as a function of sharing activity locations and sharing socio-economic and life-style characteristics. This is more difficult to derive, but the activity-based approach brings the problem closer to a solution. Once $\Im$ is established, the system merges sets $\boldsymbol{K}$ across the actors connected through the network. Note that by its communicative nature, the social network is a potentially very rich source of information in the system as well as reality.

The previous components determine the moment when a household becomes engaged in active search and the consideration set of houses on offer resulting from it. The residential choice model determines the probability of choosing $i \in \boldsymbol{I}$ in two steps:

$$\Pr(i|I) = \Pr(\max_{i \in I}\{U(X_i)\} > U^{stay})\,\Pr\big(U(X_i) > U(X_j), \forall j \neq i\big), \tag{5}$$

where $\boldsymbol{X}_i$ is a vector of attributes of house on offer i, $U(\bullet)$ is a utility function and U^{stay} is the utility of not moving. The first term on the RHS represents the probability of a positive move decision and the second term the probability of choosing i. Note that the utility of moving may not be equal to $(\max_{i \in I}\{ U(\mathbf{X}_i)\})$as the former is based on expectations and the latter on evaluation of actual houses on offer.

An example of a possible implementation is a nested-logit model:

$$\Pr(i|\mathbf{I}) = \frac{\exp\left(\frac{1}{\mu}\ln\sum_i \mu V_i\right)}{\exp\left(\frac{1}{\mu}\ln\sum_i \mu V_i\right) + \exp(V^{stay})} \frac{\exp(V_i)}{\sum_i \exp(V_i)} \tag{6}$$

or:

$$\Pr(i|\mathbf{I}) = \frac{\exp\left(V_i + \frac{1}{\mu}\ln\sum_i \mu V_i\right)}{\exp\left(\sum_i V_i + \frac{1}{\mu}\ln\sum_i \mu V_i\right) + \exp\left(\sum_i V_i + V^{stay}\right)} \tag{7}$$

where V are the structural components of U and μ is a scale parameter to be estimated. However, given the other agent in Albatross, decision tables which match residential preferences against the characteristic of the vacant dwellings are more appealing. Moreover, decision tables easily represent thresholds, substitutions and veto criteria, which are difficult to incorporate in algebraic, utility-maximising models.

The choice of attributes $\boldsymbol{X}_i$ is critical and include dwelling attributes, neighbourhood characteristics, relative location vis-à-vis nodes of the multimodal transport network, vis-à-vis work/school location, vis-à-vis centres for shopping/recreation/leisure, vis-à-vis nodes of the social network, and the social structure of neighbourhood. Given a classification of households based on typical activity-programs, household-type specific parameters determine the relative importance associated with the location attributes and, therefore, the compromise the household is willing to make regarding the activity program. Social-network and social-structure constructs also play an important role in the dynamics of the system. Note that the integration of land use and transport thus goes beyond simply treating the calculation travel times in the transport model as input to the residential choice model.

Competition between searching households is a final factor that is taken into account in allocating households to vacant dwellings. Collectively, the previous steps determine the set of candidate households for each specific house on offer. Under perfect market conditions, the price mechanism would bring demand and supply together. In n-to-1 markets (n demanders, 1 supplier), a bidding process would settle equilibrium price. However, at least the Dutch housing market is far from perfect in an economic sense. In the social sector the market is regulated, whereas in the free sector a “first one first considered” rule tends to dominate. Willingness to accept a price (by the demander) or a bid (by the supplier) is typically influenced by urgency of a purchase. Therefore, an imperfect bidding-process model is developed for this part of the system.

5 Conclusions and Discussion

This chapter has discussed the residential choice component in the Albatross and Ramblas model systems. The discussion should have made it clear that at the present stage of development, both models are primarily activity-based models of transport demand, and not integrated land use – transport models. Their prime goal is to predict activity–travel patterns and associated traffic flows. The distribution of residential land use, in terms of households and persons, is exogenously given. Based on the available data sources, a set of tools has however been developed to create synthetic populations that serve as input to the models.

Having said that, work is on its way to further elaborate these models and predict dynamic residential choice behaviour. Much of this work can be based on previous work of the authors and their co-workers. Most progress to date in terms of actual software development has been completed in the context of Ramblas. It contains a module for modelling residential choice behaviour that is used to predict the choice of residential zone for people moving house and newcomers in the housing market. Simultaneously, the properties of the dwelling stock are updated. Residential preferences measured in the National Housing Survey are matched against vacant dwellings in the market. These preferences are measured using a compositional stated preference approach, but alternatively any conjoint preference approach, mentioned in the introduction, could be used in principle. To capture the heterogeneity in residential preferences, estimated utility functions should be segment-specific or the micro-simulation should incorporate the inherent heterogeneity.

Work in progress as part of the Albatross system uses this information plus information about pressure in individuals' activity–travel patterns, and a set of other events to simulate dynamic residential choice. It represents an attempt to replace the rather rigorous assumptions underlying the utility-maximising and welfare-maximising multinomial and nested logit models by a computational process model that is based on imperfect and limited information, spatial and non-spatial search in dynamic housing markets and a suboptimal market clearing process in non-equilibrium.

While we argue that this development is theoretically appealing, it does not necessarily result in improved prediction. In that regard, the extreme differences between the data-driven Ramblas model and the theory-driven Albatross model would make a comparison of the predictive performance of these models very interesting.

References

Arentze TA, Timmermans HJP (2000) Albatross: a learning based transportation oriented simulation system. European Institute of Retailing and Services Studies, Eindhoven

Arentze TA, Timmermans HJP (2003) Albatross 2.0. European Institute of Retailing and Services Studies, Eindhoven

Louviere JJ (1984) Hierarchical information integration: a new method for the design and analysis of complex multiattribute judgement problems. Adv Consum Res 11:148–155

Louviere JJ, Timmermans HJP (1990) Hierarchical information integration applied to residential choice behaviour. Geogr Anal 22:127–145

Molin E, Timmermans HJP (2003) Accessibility considerations in residential choice decisions: accumulated evidence from the Benelux. In: Proceedings 82nd Annual Meeting of the Transportation Research Board, Washington, D.C, 12–16 January 2003

Molin EJE, Oppewal H, Timmermans HJP (1999) Group-based versus individual-based conjoint preference models of residential preferences: a comparative test. Environ Plann A 31:1935–1947

Oppewal H, Louviere JJ, Timmermans HJP (1994) Modelling hierarchical conjoint processes with integrated choice experiments. J Mark Res 31:92–105

Timmermans HJP (1989) Een decompositioneel hierarchisch model voor woningkeuze: theorie en illustratie. In: Musterd J (ed) Methoden voor Woning en Woonmilieubehoefte Onderzoek. SISWO, Amsterdam, pp 46–72

Timmermans HJP, van Noortwijk LE (1995) Context dependencies in housing choice behaviour. Environ Plann A 27:181–192

Timmermans HJP, Borgers AWJ, van Dijk J, Oppewal H (1992) Residential choice behaviour of dual earner households: a decompositional joint choice model. Environ Plann A 24:517–533

Timmermans HJP, Molin EJE, van Noortwijk LE (1994) Housing choice processes: stated versus revealed modelling approaches. Neth J Housing Built Environ 9:215–227

Timmermans HJP, Noortwijk IE, Oppewal H, van der Waerden PJHJ (1996) Modelling constrained choice behaviour in regulated housing markets by means of discrete choice experiments and universal logit models: an application to the residential choice behavior of divorcees. Environ Plann A 28:1095–1112

Timmermans HJP, Arentze TA, Dijst M, Dugundji E, Joh C-H, Kapoen LL, Krijgsman S, Maat K, Veldhuisen KJ (2002) Amadeus: a framework for developing a dynamic multi-agent, multi-period activity-based micro-simulation model of travel demand. In: Proceedings of the TRB Annual Meeting, Washington, D.C

Van de Vijvere Y, Oppewal H, Timmermans HJP (1997) The validity of hierarchical information integration choice experiments to model residential preference and choice. Geogr Anal 30:254–272

Veldhuisen KJ, Kapoen LL (1977) Een Regionaal Locatie Model. Ph.D. thesis, University of Utrecht

Veldhuisen KJ, Timmermans HJP (1984) Specification of individual residential utility functions: a comparative analysis of three measurement procedures. Environ Plann A 16:1573–1582

A Microsimulation Model of Household Location

Olga Feldman, Roger Mackett, Emma Richmond, David Simmonds and Vassilis Zachariadis

Abstract This chapter describes the development of a new microsimulation model of individual and household changes and choices within a land-use/transport interaction modelling structure.

The major strength of the model is naturally its disaggregate and dynamic nature, which means that the user can aggregate the output at any desired level of household or person characteristics, and that it is possible to trace individuals, households, jobs and dwellings over time so as to observe the modelled processes of change at a level of detail that is simply not possible in other types of models.

1 Introduction

This chapter describes the development of a new microsimulation model of individual and household changes and choices within a land-use/transport interaction modelling structure. The work was carried out by David Simmonds Consultancy (DSC) in collaboration with MVA Consultancy, University College London and the University of Leeds School of Geography, under a commission from the UK Department for Transport (DfT). Further developments are being considered, so inevitably the chapter is limited to describing the model at a particular point in time.

O. Feldman (✉), E. Richmond, and D. Simmonds
David Simmonds Consultancy, 14 Millers Yard, CB2 1RQ Cambridge, UK
e-mail: OlgaFeldman@tfl.gov.uk

R. Mackett
Centre for Transport Studies, University College, Gower street, London WC1E 6BT, UK

V. Zachariadis
David Simmonds Consultancy, 14 Millers Yard, CB2 1RQ Cambridge, UK
Department of Architecture, University of Cambridge, 1-5 Scroope Terrace, Trumpington Street, Cambridge CB2 1PX, UK

F. Pagliara et al. (eds.), *Residential Location Choice*, Advances in Spatial Science,
DOI 10.1007/978-3-642-12788-5_11,

The Chapter is organized as follows. In subsequent sections the following aspects of the model are discussed. Section 2 discusses the choice of modelling approach. Section 3 describes the area covered by SWYSimM.

Section 4 focuses on the aggregate components of SimDELTA. Section 5 describes the static components of the microsimulation modelling, that is, the creation of the initial microdata for the base year, and the sources used in this, whilst Sect. 6 describes the overall structure of the dynamic model and the design of the microsimulation components. Section 7 draws some conclusions on the status and possible applications and future developments of the model.

2 The Choice of Modelling Approach

The decision to develop a microsimulation-based model arose both from the DfT Specification and our own thinking, in particular our earlier "New Look at Multi-Modal Modelling" for the Department (Simmonds et al. 2001). The general arguments in favour of highly disaggregate modelling are well established. There is however a continuing debate in many areas about the relative merits of what is sometimes called "econometric" disaggregate modelling, on the one hand, and strict microsimulation on the other (see, for example, Bowman and Ben-Akiva 1997). Both techniques work on samples of individual decision-makers (persons or households). The essential difference is seen in their modelling of choices between discrete alternatives (e.g. between modes, or between residential zones):

- In the "econometric" approach, each modelled decision maker will have a non-zero probability of choosing each available alternative, and these probabilities are used directly as the results of the model – so each modelled individual is assumed to spread out across the available alternatives for each choice.
- In the microsimulation process, whilst the same probabilities may be calculated (sometimes in exactly the same way), each modelled individual is allocated to one single alternative.

Another important difference is that microsimulation often uses Monte Carlo simulation methods. In this case, random numbers are used in the process of "deciding" which of the available alternatives the decision-maker will choose, given the calculated probabilities. This means that if the model is rerun with different random numbers, the results of the model will be different. This raises a number of issues about the practice of using such a model, and about the interpretation of the results. (Note that where the microsimulation works on deterministic rules, these issues do not arise, except in so far as the base data is itself typically a microsimulation output with a random component.)

"Econometric" approaches to disaggregate modelling have been used extensively in transport modelling, and were identified in the "New Look" work as representative of the state-of-the-art techniques for application of the conventional "four-stage model" approach. Microsimulation approaches are used extensively in traffic modelling, especially at the most detailed levels, though here too there are issues about the use and interpretation of Monte Carlo results (see Feldman and Maher 2004).

Household location modelling has for some time tended to move towards micro-simulation rather than econometric disaggregate modelling. The key examples are IRPUD (Wegener 1982), MASTER (Mackett 1990, 1992, 1993), UrbanSim (Waddell 1998; Waddell et al. 2003) and the TLUMIP model of Oregon (see chapter, "Stated Preference Examination of Factors Influencing Residential Attraction" by Hunt). The reasons for this trend to microsimulation are:

- The possibility of relating to the range of other microsimulation work on household and individual change over time.
- The problems that arise, both conceptually and practically, with econometric models where a moving household will be distributed in small fractions across many locations – it is much easier to design and build a model where one household moves from one initial location to one new location.
- The possibility of building explicit consideration of available information, information-gathering and search processes into microsimulation; most practical forms of econometric choice modelling (i.e. logit models) assume perfect knowledge of all available alternatives.
- The availability of increased computing power.

Given this background, and the requirement to focus on modelling households and individuals (rather than employment, firms or development processes), our approach to the development of this model was that

- The overall structure would remain that of our existing DELTA models (see chapter, "The DELTA Residential Location Model"), but with the household/person processes being rebuilt as microsimulation models, exploiting the modular structure of DELTA.
- The microsimulation components should explicitly model changes to members of the sample over time (rather than, as in many other microsimulation models, generating a separate sample for each modelled period of the forecast).
- The microsimulation modelling would be carried out at ward level (see below).
- The default designs for the additional elements of microsimulation modelling would be based on those from the earlier MASTER model (making use of lessons learned from the MASTER projects).
- The model would be tested on the South and West Yorkshire areas, making use of existing DELTA and transport models there so as not to have to build these anew.
- The emphasis would be on getting a working model up and running, and on identifying the needs for further research (and possibly new surveys) from the model implementation and testing processes.

The decision to build the model was important. Wards in England are small electoral and statistical units, those in South and West Yorkshire having an average of about 5,000 households each. The pre-existing land-use model of the area was implemented using zones which were generally groups of wards. The new model was therefore designed to operate on a rather finer spatial system but not to operate at a micro-spatial (parcel or fine grid cell) level. Hence, although the model operates on lists of dwellings, the location of dwellings is not currently defined below the ward level.

The term "microsimulation modelling" generally covers all of the possibilities of:

- The processes of generating a micro-level (individual household and/or person) data consistent with a given aggregate data set (whether observed, as in Census tables, or forecast by an aggregate method such as cohort-survival population forecasting)
- Modelling the impacts of changes by analysis within such a micro-level dataset (e.g. to look at the impacts of job losses/gains on the microsimulated population at a particular point in time) and
- Modelling changes over time by applying microsimulation techniques to the processes of change at the household/person level

It is worth emphasising at this point that the objective of the present project was to develop a fully dynamic microsimulation model in which all of the processes affecting households and their members would be modelled as changes over time – i.e. the focus is on the last of the possibilities listed above. The project therefore split into two main stages: the generation of the initial sample of households and household members (a process referred to as the static model), and the implementation of the dynamic model proper, forecasting changes over time.

The new modelling package has been given the name SimDELTA. The choice of the South and West Yorkshire case study area for the initial application of SimDELTA allowed the modelling to be based on the existing South and West Yorkshire Strategic Model (SWYSM – see Simmonds and Skinner 2004; Feldman et al. 2007). SWYSM was originally developed for the South and West Yorkshire Multi-Modal Study; that study also involved the building of more detailed, generally ward-level, highway and public transport models which were used to provide transport inputs to the new microsimulation modelling. This first application of SimDELTA has been named SWYSimM.

3 The SWYSimM Area

The SWYSimM area is a subset of that modelled in SWYSM (see Fig. 1). The definition of the SWYSimM area took account of the work on Functional Areas and Regions which MVA Consultancy and DSC carried out for DfT in an earlier preparatory study (Feldman et al. 2005b). The Functional Area analysis provided a catalogue of possible zoning systems at different levels of aggregation. At any chosen level, the Functional Areas are areas of relatively high self-containment in the travel-to-work patterns. The SWYSimM application covers five large areas covering virtually all of the South and West Yorkshire areas and some adjoining

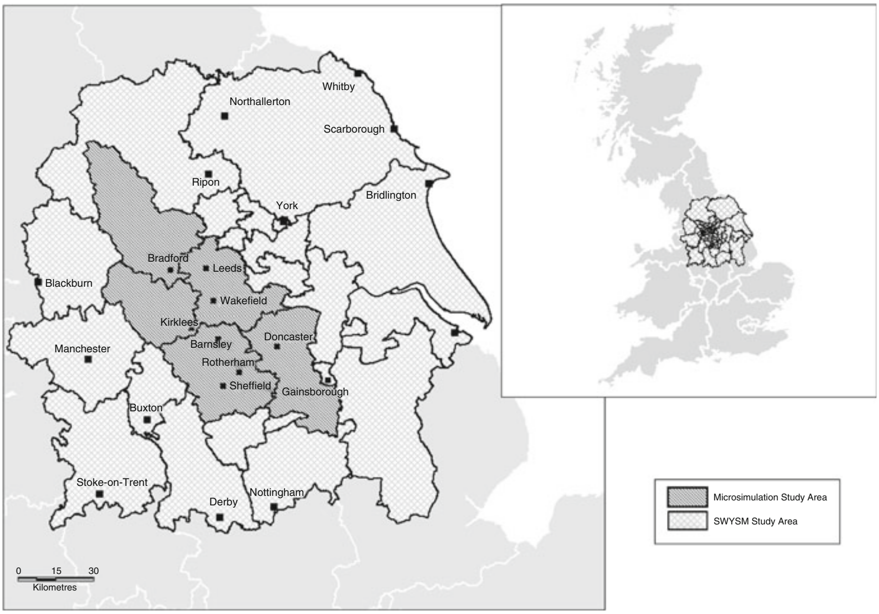


Fig. 1 SimDELTA study area

territory; these are represented at ward level, giving 283 microsimulation zones. The rest of the SWYSM modelled area forms a set of 18 external zones for SWYSimM. These cover the rest of the South and West Yorkshire plus the larger adjoining areas of Greater Manchester, Humberside, Lincoln, Nottingham, Derby, Stoke, East Lancashire and Yorkshire. In these external zones population is still modelled at the aggregate rather than microsimulation level using the standard DELTA package.

4 The Aggregate Components of SimDELTA

Before launching into the discussion of the microsimulation components of SimDELTA, it is useful just to identify the parts of the overall model which remain at the aggregate level. These are

- The transport modelling
- The modelling of longer-distance migration
- The modelling of employment location and
- The modelling of non-residential development

The interface from the microsimulation model to the aggregate level is a simple one of aggregating microsimulation outputs and superimposing them on the aggregate model. The interface from the aggregate sub-models to the microsimulation involves:

- Calculating zonal accessibilities (from the transport model outputs and the land-use forecasts) by microsimulation zone.
- Converting the aggregate migration model outputs into probabilities to be applied by Monte Carlo simulation at appropriate stages in the microsimulation modelling.
- Converting the aggregate employment forecasts into redundancy rates and/or synthetic microdata representing new jobs.

The microsimulation is inserted into the standard DELTA sequence as shown in Fig. 2. Note that the intention is that the forecasting of housing development should also remain in the aggregate modelling, which means that an interface would be needed to turn changes in dwellings stock into microdata on new dwellings (or into instructions to demolish a proportion of existing dwellings). This has not yet been

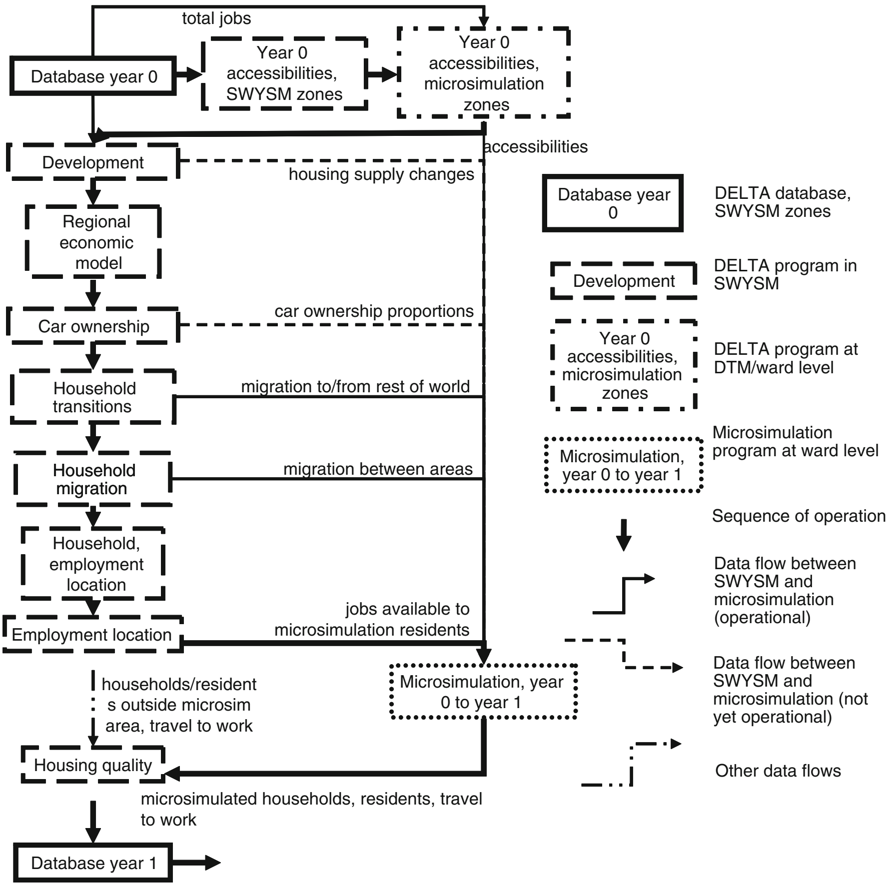


Fig. 2 SimDELTA: linkages between aggregate and microsimulation components in 1 year Note that only the data flows affected by the addition of the microsimulation process are shown; there are numerous other flows between DELTA components and databases

implemented, and the SWYSimM runs to date have all been undertaken with exogenously prepared micro-level inputs for changes in housing stock.

5 Static Model Implementation

Whilst the focus of the project was on the modelling of household changes over time, i.e. the dynamic modelling, the initial static modelling was an essential part of the work programme because, as usual, it was necessary to generate a dataset from which the dynamic modelling would operate. The static modelling itself involved two stages. The first was the generation of an initial synthetic microdata set, based on data from the regional Sample of Anonymised (household) Records (SARs), to produce a synthetic 100% sample of households and persons for each zone (ward) closely matching the characteristics of the zone's population as shown by the published (aggregate) Census tables. The second stage involved the addition of further variables to this data; this was necessary partly as a result of the process of synthesizing zonal data from a regional sample (even if the SARs contained data on individuals' workplaces, this data would not be valid for the synthetic data based upon the SARs sample), and partly because some required variables were not available in the SARs data at all.

The initial generation of the synthetic microdata was carried out by a combinatorial optimisation method called simulated annealing, which has been widely used in other static spatial microsimulation applications (see for example Ballas 2001; Ballas et al. 2004; Williamson et al. 1998). The starting point was the simulated annealing modelling method used in ULSG's SimLeeds project (Ballas 2001). This was applied so as to select household records (with repetition) from the relevant regional sub-set of the 1991 Household SARs which would match the observed population of each zone as reported in the 1991 Small Area Statistics tables. Both sets of data were derived from the 1991 UK Census of Population.

The simulated annealing procedure in the SimDELTA context can be summarised as involving the following steps, applied independently for each zone (ward):

- Taking a random sample of N microdata household records from the overall set of microdata, by sampling with replacement, where N is the number of households in the zone.
- Tabulating the characteristics of the sampled microdata.
- Comparing these with the chosen tables of observed Census data for the zone and assessing the goodness of fit of the sample data to the observed data.
- Randomly replacing some of the cases in the sampled microdata and repeating the above two steps.

The last three steps are repeated until a sample of microdata is found for the zone which produces a satisfactory match to the tables of observed data for the zone. If any replacement of cases results in a significantly worse goodness of fit, the replacement is generally reversed and a different random replacement is attempted. As the simulation progresses and (hopefully) the goodness-of-fit improves, the number of records selected for swapping at one time decreases. This allows faster change early in the process, whilst proceeding more cautiously once the fit has improved significantly. The static model also employs a restart method which is applied if the model fails to find a satisfactory solution within the maximum permitted iterations; in this case, the simulated annealing process begins again with a wholly new initial sample of records. The simulation is complete when the total relative error is less than a specified target. For further detail of the simulated annealing process, see Feldman et al. (2005a).

Two key points should be noted about the initial generation of the synthetic data. The first is that the simulated annealing process is itself a microsimulation model with a highly significant random element, and hence the synthetic population that results is probably only one of many possible populations which could be generated with similar levels of goodness of fit to the observed data. The amount of computing necessary to produce just one synthetic population meant that has not yet been possible to explore the consequences of working with different but equally appropriate populations. The potential for detectable variation amongst such possible populations depends in part on the number of different variables in the datasets and the number of these variables which are considered in considering the goodness of fit resulting from the simulated annealing. This leads to the second point, which is that the simulated annealing process can only practically test goodness of fit against a few out of the dozens of univariate or bivariate tables available for each ward in SAS. The present exercise used ten tables, covering many but not all of the possible dimensions of the data; different tables were given different weightings in assessing goodness of fit.

The second stage of the static modelling adds

- The socio-economic group of persons of potential working age or above who do not have a socio-economic group in the SARs data.
- Further detail of economic status, if this is insufficiently defined in the SARs data.
- Whether the individual holds a (car) driving licence.
- The workplace and wage or salary of each working individual in his/her current job (if any).
- Household income (taking account of individual earnings and of benefits, pensions etc).

All of these are implemented using Monte Carlo simulation to apply appropriate distributions. In the case of the workplace, the distribution is taken from the Census Travel-to-Work tables, which are available at ward level. This second stage of the static modelling also:

- Identifies which households consist of persons sharing (i.e. have no other relationship that would cause them to wish to live together) and which of these consist of students at their term-time addresses and

- Produces the initial list of single/separated/widowed/divorced women which constitutes the set of potential partners for the couple-formation modelling in the first year of the dynamic model

All of the variables set up in the static modelling are modified over time within the dynamic model. The static modelling is therefore used only in the base year.

6 Dynamic Model Implementation

The output of the static microsimulation model provides the input for the dynamic microsimulation model, which has been run from 1991 to 2011. Much of the work concentrated on the period 1991–2001 in order to compare results with data from the 2001 Census of Population.

The overall structure of the model is broadly similar to the other microsimulation models mentioned earlier in that it starts with demographic changes to individuals, then deals with changes in household composition, and finally with employment and household location/relocation, on the assumption that the later processes are generally more dependent on the earlier ones. This structure is illustrated in Fig. 3. Note that except for the couple-formation element of the household composition stage, the model can be run for one household at a time.

6.1 Individual Demographics and Other Changes

Ageing. The aging process is straightforward. The age of each person is increased by 1 each year.

Survival. The probability of an individual surviving the year is a function of age and gender, based on official actuarial statistics.

Birth/multiple birth. Births are modelled using birth rates by 5-year age group, ethnicity, and the mother's couple status. There is a constant probability that a birth will produce twins (the possibility of triplets or more is ignored). The gender of the child is fixed using probabilities for the ratio of males to females. The attributes of a new-born child are set as follows: age is zero, sex is determined probabilistically, couple status is single, ethnicity and location are those of the mother. All the other personal characteristics are undefined. In the next simulation period, the new individual is simulated along with the other individuals in the household.

Socio-economic status. All adult persons within the model were assigned one of the four socio-economic groups aggregated from the greater detail in the Census, namely,

- Seg1 – professional and managerial.
- Seg2 – junior professional, non-manual supervisor, etc.

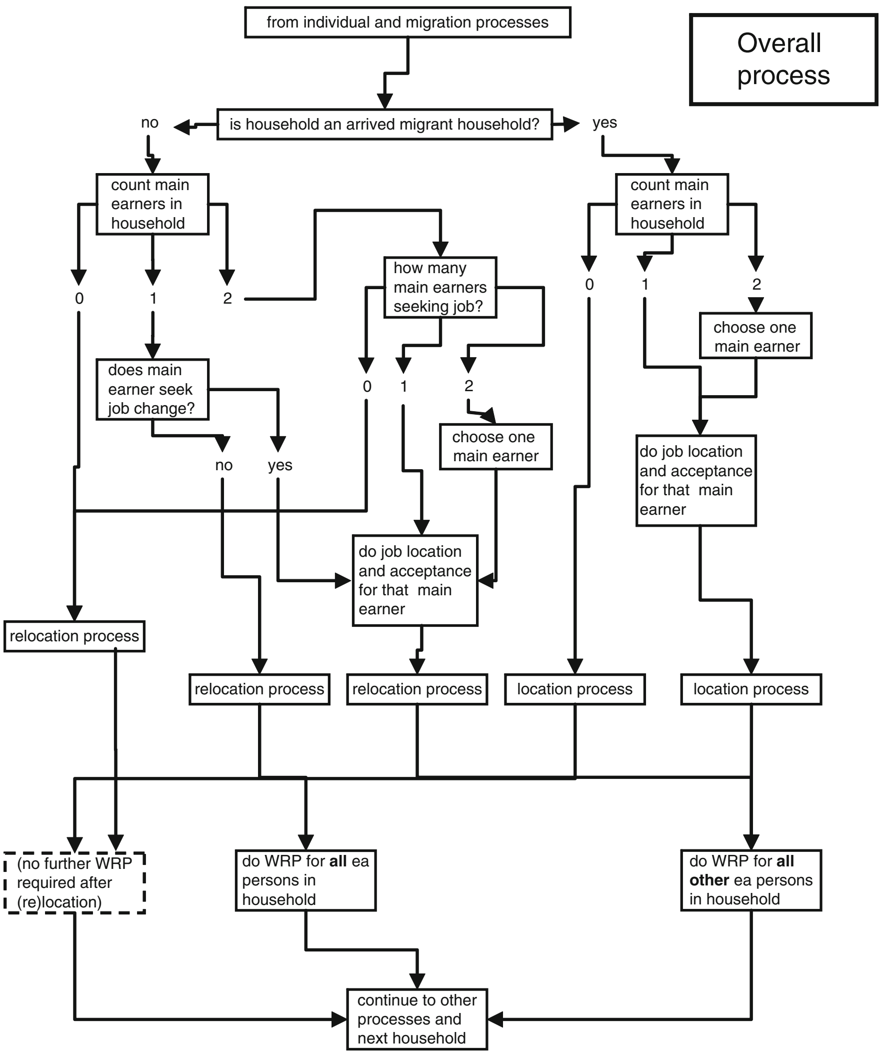


Fig. 3 Overall process of residential/work choices (WRP – work related processes)

- Seg3 – skilled manual.
- Seg4 – semi-skilled and non-skilled manual.

This was based on their SARs socio-economic group where applicable or generated using Monte Carlo simulation if their SARs socio-economic group was "not applicable" or "not adequately described" Every child is automatically attributed the socio-economic group 0 – non-defined until reaching working age (16 years). On completing education every college graduate enters the market with socio-economic group 2 while any other person enters the market with socio-economic group 4.

We use Monte Carlo simulation to allow people to move to other socio-economic groups throughout the rest of their working life. The probabilities in this process depend on age, sex and their current socio-economic group. Persons search for employment compatible with their socio-economic groups; if this is not available they expand their search accordingly. Applicant acceptance is also conditional by the job/worker socio-economic group match.

Educational status. Education attributes in the UK census data are based on "level of highest qualification" values for persons over 16 years old. Persons under 16 have no relevant education attribute. Information for newly processed persons is accordingly updated in the model based on his/her education status (student at 18, student at 21, etc.). The model also allows persons to become students at older ages.

Entering/re-entering labour market. Each person in the model is assigned one of 11 possible economic statuses (not applicable, employee full time, employee part time, self-employed with employees, self-employed without employees, government scheme, unemployed, student, permanently sick, retired, other inactive). Once the person is employed, he or she normally stays employed but can change job, become retired, become permanently sick, become redundant or become "other inactive". To address withdrawal from the market due to family related matters, mothers have a probability of becoming "other inactive" after the birth of a child; they then have a probability of re-entering the labour market based on their socio-economic group and the age of their youngest child. Other persons whose economic status is inactive (but not students) have a fixed probability of re-entering the labour market.

Redundancy: Within an apparently static situation, jobs are usually being lost due to the decline/closure of individual firms/establishments, whilst an equal number of similar jobs are being created due to the growth of other firms/establishments in the same zone and sector. To represent the effects of job losses, the model has to apply redundancy probabilities which are calculated in the aggregate microsimulation interface. If a worker is made redundant, he/she will not be able to seek another job in the same year (unless he/she changes household and the new household is considered later that year).

Retiring from labour market: Probability of retirement is defined as a function of age and gender, with most men retiring around age 65 and women around 60. At age 75 all people who are still employed retire. For each worker who retires the number of vacant jobs in the model increases by one. In principle, some people who are retired may choose to re-enter the labour market but we are not considering such movements.

Acquiring/losing driving licence. Probabilities of obtaining or losing a driving licence (according to age and sex) are applied to all people over 17.

Becoming permanently sick. Any economically active person may at any time become permanently sick. In this case the person leaves his or her job and the job market permanently. The probabilities of becoming permanently sick are based on age, sex and socio-economic group.

Moving to institution. Persons aged 65 and over have a probability of moving into institutional accommodation. In some cases, this is a temporary move; the model allows for a small proportion to return to their previous dwelling within 1 year; the remainder are assumed to remain within institutional accommodation for the rest of their lives. They therefore leave the household population; if they were living alone, their dwelling becomes available to another household.

6.2 Household Changes

Couple formation and marriage: A "male-dominant" model is implemented whereby "couple formation" is treated as a male choosing a partner from a list of eligible females. The probabilities of the male seeking a partner, and of a female being eligible, are input as a function of age and existing couple status. A new couple can be married or be cohabiting: cohabiting couples may later marry. We assume that partners are usually found within the area of residence however there is a low probability that partners are found in different locations. The model allows for migration of people who do form couples with persons from other regions such that one of the partners moves to the location of the other partner. Monte Carlo simulation is used to identify who is moving. The model does not so far form new same-sex couples.

Separation: The divorce and separation of married and cohabiting couples is modelled using probabilities based on age and whether married or cohabiting.

Absence from households: There are probabilities that young people will leave their parents' home (e.g. to study) and then come back after a number of years.

Student only and shared households: The model allows for unrelated people to form shared households, and for these households to dissolve or reform over time.

Obtaining/losing car: The SWYSM land use model has a car ownership sub-model which works entirely in terms of the zonal probabilities of a household of a particular type owning no car, one car or two-plus cars. These probabilities are input to SWYSimM and used to generate the probabilities for individual households of acquiring an additional car or giving up a car.

Household income: The household income is simulated as the sum of the incomes of each member of the household. Working persons contribute their wages, while the unemployed, retired and permanently sick contribute pensions and benefits. Children contribute through benefits and tax policies. Mature students involved in further education are considered to retain the wage of their last job.

Housing affordability: Households form budgets for buying or renting based on their tenure preferences, the values of the housing markets and the characteristics of available housing stock. In case of renting, household budgets are formed as a proportion of the household gross income which varies between 25 and 35%. Buying budgets are based on a number of parameters including savings, net household income, outstanding mortgages from previous acquisitions and previous type of tenure. Savings are calculated each year for every household after subtracting costs of living (transport costs, foods and goods costs, taxing costs etc) and housing costs (rent or mortgage) from annual housing income. Outstanding mortgages are passed on each year after subtracting a sum equal to a set proportion of the household annual income. The buying budget is formed by adding savings and – in case of owning one – the current dwelling's value after subtracting outstanding mortgages. Households also add to their budgets a new mortgage based on their current net income and the current multiplying factor. In the case of outstanding mortgage the value of the mortgage is decreased. Students and non earner households are not eligible to get a new mortgage, while sharer and student households cannot change their tenure preference to owner-occupied.

6.3 Employment

Job supply. The job supply, in terms of full and part time jobs, for the people living in the microsimulation Study Area and working in the microsimulation Study Area or in the microsimulation external zones are output from the SWYSM model. Jobs contain attributes such as current wage and location.

Main earner: We have introduced this concept as the main earner, or earners, in a household could have a strong influence on household decision making especially with respect to location. The main earner is the person who is likely to be the main income-earner of the household – although he or she may be unemployed at a particular point in time. The new definition is intended to avoid problems with the inconsistency of "head of households" and "household reference person" in the 1991 and 2001 UK Censuses. We assume that in any household with just one economically active person that person is the main earner. If a households contains a couple (married or cohabiting) who are both economically active, of equal socio-economic group and in white-collar or skilled manual level occupation, we assume this to be a dual-career household and that the members of the couple are joint main earners.

In all other households containing more than one economically active person, we take the oldest person in or seeking full-time employment as the main earner. If no one is in or seeking full-time employment, we simply take the oldest person.

Job and workplace choices: Unemployed persons, those entering the labour market for the first time and those returning to it after a break (maternity leave) look for a job in areas sorted according to the areas' utilities. Utility is calculated according to the generalised cost to travel from the current area of residence and the number of suitable available jobs in that area. Each agent "applies" for a specified number of jobs in each area and searches only in a specified number of areas. Unemployed persons unsuccessful in finding a workplace that suits their preferences, are likely to decide to search for jobs of other socioeconomic groups or other economic types (part-time instead of full-time and vice versa). If an application is accepted the unemployed agent accepts and stops job seeking.

Seek-to-change-job process: The process followed by an agent that already occupies a job is fairly similar to the one for the unemployed and agent's entering/re-entering the labour market. However in this case areas are also sorted according to the distance from the current work location. After a job application has been accepted, the employed agent compares the proposed wage to their current. If the wage is lower the new job is rejected.

Accept/reject candidate, accept/reject job: Probabilities are applied to decide whether the job is offered to the potential worker and whether it is accepted. The probabilities depending on how closely their profile matches the attributes of the job.

Wages: Each working person is assigned a specified annual wage based on their characteristics (e.g. age, socio-economic group, gender etc) and the job they occupy.

6.4 Household Location

Housing stock: The change in the number of dwellings over the period 1991–2001 is an exogenous input to SWYSimM model, as it was to SWYSM. For later years the changes in housing stock are forecast within DELTA.

Each dwelling is assigned a tenure depending on whether it is owner occupied, privately rented or rented from a local authority or housing association. All dwellings continue to be of the same tenure as in the initial database or retain the same tenure they were initially assigned.

Four dwelling types are modelled: detached, semi-detached, terraced and flat. In the synthetic database each dwelling belongs to one of these four types (from SARS 1991). The dwellings also differ by the dwelling sizes (number of rooms). Currently we do not have information on the number of new dwellings built by dwelling type and the number of rooms, so a "cloning" process is used: a dwelling in the current database is randomly chosen and its characteristics are copied to the new dwelling.

Seek to move: After budget formation, households seeking to move search within their preferred areas for a vacant house matching their preferences (tenure and budget) and sized within their size tolerance (usually one room tolerance). In order to avoid futile searches, households seeking relocation check whether their budget is over the expected minimum budget for a house of the required size in an area before searching it. Households search a fixed number of appropriate areas and if they are unsuccessful in finding a property, look for alternative tenure types before giving up. Unsuccessful external in-migrants are deleted, i.e. assumed not to migrate into the modelled area.

If a household finds a suitable vacant dwelling it marks it as a potential target. In the event that a household finds more than one suitable dwelling it always prefers the one closer to its budget in order to maximise utility. Areas are sorted based on their utility based on area deprivation, distance for current area of residence, general accessibility of target area and generalised cost of target area to main-earners work place.

Housing tenure choice: Household's choice of tenure is influenced by the supply of dwellings of each tenure type. Households unable to find accommodation of their preferred tenure within their budget constraints can switch to a different tenure type.

Dwelling choice. The dwelling that is chosen by a moving household or individual must fit the required characteristics. Households moving due to high room stress (too many people per room) can only move to a more suitable (lower room stress) property. If none are available they have to wait.

Housing prices or rents: The household location model requires "asking prices" to be set for owner-occupied dwellings which are being sold, and "asking rents" for dwellings which are being let. The rents are modelled as fractions of the sale price. The sale price is calculated using a hedonic price model based on:

- The price of a typical dwelling of a particular type in this zone (this data is available to 2005, beyond 2005 the 2005 price is used and the inflation index is applied).
- The location constant of this zone in this year.
- The cumulative inflation rate of housing prices from the base year.
- The average price of a room according to dwelling type in the base year.
- The size difference in terms of number of rooms from the average for each dwelling type.
- The market change indicator which reflects demand and supply.

Location choice: A price or rent-based location model is implemented. Households trade off desirable housing, location and accessibility characteristics against price or rent, and price or rent are adjusted over time in response to changes in the balance between supply and demand.

Household location/relocation and migration: The overall model sequence is presented in Fig. 3. Job choice for main earners may occur before residential relocation (i.e. change of job by the main earner can lead to household relocation); job choice for others (and possibly for main earners) is considered after household relocation (in the next year).

Before considering either change of job or change of dwelling, we test whether the household is going to migrate (make a longer-distance relocation necessarily involving a change of job(s)). If so, then they disappear from their existing area – and may reappear elsewhere in the model as migrants into another area.

For households which have not migrated, we consider possible job changes by main earners (which could have a strong influence on household decision of whether/where to relocate), household relocation, and possible job changes by other household members.

For households with one main earner, we first test whether he/she is seeking a new job (i.e. whether he/she is employed and seeking to change job, or is currently unemployed). If so then we model work-related choices for that person and then we model relocation choices for the household, which will be influenced by the job location. Then (whether or not relocation results) we model work-related choices for any other workers or unemployed persons in the household. If the main earner is not seeking a new job, we model location choices. If relocation occurs, we model job choices for all working members of the household; if relocation does not occur then we model whether-to-change job for the other working members of the household (if any other than main earner exist).

For households with two main earners, i.e. dual career households, we

- For each of the main earners test whether he/she is seeking a new job (i.e. whether he/she is employed and seeking to change job, or is currently unemployed).
- If both are seeking a new job, then randomly choose one of them to make work-related choices before relocation choices, allowing the other one to make work-related choices afterwards (whether or not relocation occurs).
- If only one is seeking a new job, he or she is limited to doing this after relocation choices have been considered.

As with single-main-earner households, main earners in two-main-earner households who do not seek a change of job before considering relocation but who do then relocate are tested next year again to see if they then wish to seek a change of job. (Those who do not relocate are not tested again.)

6.5 Completion of the Microsimulation

The final step of the dynamic microsimulation is to generate output to update data for further analysis, for the production of summary outputs to pass to the aggregate (DELTA) components of SimDELTA, and as the starting point for the dynamic microsimulation model in the following year. Some summary information on changes to households and persons is produced as a matter of routine, but since households, persons and dwellings all have individual identifiers, further longitudinal analysis can readily be carried out by merging datasets for different years using standard software such as SPSS.

7 Conclusions

The present situation (July 2007) is that the SWYSimM application of SimDELTA is operational and producing reasonable results overall. Further work is ongoing to test it in more detail and to demonstrate its value when compared to simpler models or to complex aggregate models such as the original SWYSM. Further work is looking in particular at some of the implications of the variability of results from Monte Carlo simulation and at how this can be managed in application of the model.

The major strength of the model is naturally its disaggregate and dynamic nature, which means that the user can aggregate the output at any desired level of household or person characteristics, and that it is possible to trace individuals, households, jobs and dwellings over time so as to observe the modelled processes of change at a level of detail that is simply not possible in other types of model.

The model is potentially of substantial value as the basis for a wide variety of further work, though this has to be qualified by saying that any such possibilities would be subject to further calibration and testing of the model and of the particular model features which are most important to the application in question. The main difficulty in implementing the model is the complexity of model calibration. This involves hundreds of model parameters which often need to be calibrated simultaneously. It is very easy to get trapped in a never-ending circular process fixing one problem only to find a new one somewhere else. Therefore, as a strategy for calibration, it was essential to break the process into a series of logical, sequential steps.

The most obvious of the possibilities for further development is to continue the process of model calibration by further and more formal development of non-compensatory and rule-based location and job choice models. The formal calibration and validation of these would almost certainly require new surveys and the development of new calibration methods, or at a minimum the application of some non-conventional calibration methods. Note that the rationale for pursuing non-compensatory and/or rule-based models is not necessarily to suggest that such models should replace conventional compensatory (e.g. logit) choice models in applied modelling practice. It might well be that (apart from research application) the non-conventional models would be best used to inform other modelling by using their results as "data" to be used to calibrate conventional models. (This would amongst other things resolve the problem of how to use the multiple forecasts produced by microsimulation models – they would produce multiple datasets all of which would form multiple sets of "observations" feeding into the calibration process.)

The model could also be used to generate an artificial sample for analysis of transition/formation/dissolution patterns, complementing the limited and small sample information available from analysis of British Household Panel Survey (BHPS). BHPS has only a 5,000-household sample and has particular problems in dealing with additions to/departures from the main sample. In addition, the microsimulation model can forecast these rates for the future on the basis of a detailed, person-level demographic model. In using household transition rates as the main element of demographic modelling in DELTA, we have never claimed that the application of these rates is a sufficient demographic model in itself; we have always adjusted the rates so that the model reproduces more detailed population and household forecasts. The SimDELTA design provides a means of generating such forecasts and directly obtaining the corresponding transition rates. As in any demographic forecasting, of course, the results will be sensitive to assumptions about migration to and from the rest of the world, which, as noted earlier, are input to the aggregate modelling in SimDELTA as in the standard DELTA model.

Both the base dataset and the SimDELTA forecasts could be used as a 100% household/population sample for use in other work, e.g. as input to activity-based travel modelling or microsimulation-based car-ownership modelling. This ought to be superior to the conventional approach (in various forms of disaggregate transport modelling) of taking a base year sample population and reweighting it to match the forecast year total population, in that all of the variables should be systematically updated.

The modelling work described here has the important potential to contribute to understanding the consequences of planning policy and, potentially, to forecasting the impacts of possible future policies.

Acknowledgements The development of SimDELTA was commissioned by the UK Department for Transport. We are grateful to the original Department for Transport Project Officer, Russell Harris, his successor, Mo Shahkarami and the members of the project Steering Group for helpful discussions throughout the course of the project. All the opinions expressed are those of the authors and not necessarily those of the Department. We would also express our thanks to all of our colleagues who have contributed to the SimDELTA project or to earlier work on SWYSM. We remain responsible for all errors whether of commission or omission. The Census Small Area Statistics were provided through the Census Dissemination Unit, University of Manchester, and the SARs through the Census Microdata Unit, University of Manchester, both with the support of ESRC/JISC/DENI.

References

Ballas D (2001) A spatial microsimulation approach to local labour market policy analysis. Unpublished PhD thesis, School of Geography, University of Leeds

Ballas D, Kingston R, Stillwell J (2004) Using a spatial microsimulation decision support system for policy scenario analysis. In: van Leeuwen J, Timmermans H (eds) Recent advances in design and decision support systems in architecture and urban planning. Kluwer, Dordrecht, pp 177–192

Bowman JL, and Ben-Akiva ME (1997) Activity based travel forecasting, in Activity-Based Travel Forecasting Conference, June 2–5, 1996: Summary, Recommendations and Compendium of Papers, New Orleans, Louisiana. USDOT report #DOT-T-97–17, 32 pages

Feldman O, Maher M (2004) Optimisation in paramics: the nature of the problem. Proceedings of the 36th Annual Conference of Universities' Transport Study Group, January

Feldman O, Simmonds D, Ballas D, Clarke G, Gibson P, Jin J, Stillwell J (2005a) A spatial microsimulation approach to land-use modelling. Paper presented at CUPUM conference, 29 June–1 July 2005, London

Feldman O, Simmonds D, Troll N, Tsang F (2005b) Creation of a system of functional areas for England and Wales and for Scotland. Proceedings of the European Transport Conference, PTRC, London (on CD)

Feldman O, Simmonds D-C, Nicoll J (2007) Modelling the regional economic and employment impacts of transport policies. Paper presented to the 20th Pacific Regional Science Conference, Vancouver, May

Mackett R (1990) MASTER Model (Micro-Analytical Simulation of Transport, Employment and Residence). Report to the Transport and Road Research Laboratory on Contract 842/55, July

Mackett R (1992) Micro simulation modelling of travel and locational processes: testing and further development. Report to the Transport and Road Research Laboratory, March

Mackett R (1993) Micro simulation modelling of travel and locational processes: results. Report to the Transport and Road Research Laboratory, May

Simmonds D-C, Skinner A (2004) The South and West Yorkshire strategic model. In: Stillwell J, Clarke G (eds) Applied GIS and spatial analysis. Wiley, Chichester

Simmonds D, Bates J, May A-D (2001) A new-look at multi-modal modelling. Proceedings of the European Transport Conference, PTRC, London (published on CD)
Waddell P (1998) Development and calibration of the prototype metropolitan land use model transportation and land use model integration program, Final report on phase 2
Waddell P, Borning A, Noth M, Freier N, Becke M, Ulfarsson G (2003) Microsimulation of urban development and location choices: design and implementation of urbansim. Netw Spat Econ 3(1):43–67
Wegener M (1982) Modelling urban decline: a multilevel economic-demographic model for the Dortmund Region. Int Reg Sci Rev 7(2):217–241
Williamson P, Birkin M, Rees P (1998) The estimation of population microdata by using data from small area statistics and samples of anonymised records. Environ Plann A 30:785–816

Conclusions

Francesca Pagliara and David Simmonds

Abstract This chapter reports the models presented in the previous chapters, comparing them through the identification of some criteria. The latter are factors influencing residential choice; the treatment of dynamics; issues of interdependence and representation of planning policies or zoning controls. The final point is to stress the variety of ways in which residential location modeling (and urban modeling more generally) is advancing. It includes both continuing refinements to model packages which have a long history, and wholly new developments, and demonstrates the very different ways in which the subject is being addressed.

1 Introduction

This book has attempted to draw together a selection of recent work which is reasonably representative of the range of approaches being taken to modelling residential location within the context of developing land-use/transport interaction (LUTI) models. This concluding chapter does not attempt to draw conclusions in the conventional sense of attempting to decide what is right or wrong, good or bad, nor does it try to provide a summary description of all of the models considered in this book. All that we attempt to do is to offer some overall comments guided by the different dimensions of modelling which Professor Wilson identified in his Foreword, and some thoughts about how the development of similar models may evolve in future.

F. Pagliara (✉)
Department of Transportation Engineering, University of Naples Federico II, Via Claudio 21, Naples 80125, Italy
e-mail: fpagliar@unina.it

D. Simmonds
David Simmonds Consultancy, 14 Millers Yard, Cambridge CB2 1RQ, UK
e-mail: david.simmonds@davidsimmonds.com

F. Pagliara et al. (eds.), *Residential Location Choice*, Advances in Spatial Science,
DOI 10.1007/978-3-642-12788-5_12,

For brevity we refer to the models or projects described in the previous chapters by the short names listed in the Table 1 reported in the Chapter on "The State-of-the-art in Building Residential Location Models".

In these comments we try to keep in mind that the models described range from one-off projects to software packages with multiple and differing applications. We would emphasise again that this chapter is not intended as a summary of all the models represented in this book; hence if a particular model is not mentioned in discussion of a particular modelling characteristic or feature, it does not mean that the model does not have that characteristic or feature, nor does it imply any criticism of the model in that respect.

2 The Representation of the System

All of the models considered are explicitly spatial, as an essential condition for empirical analysis or forecasting of residential location. There is however a considerable range of spatial detail, and a distinction between the conventional zone-based models and those (UrbanSim, Oregon2) which operate on much smaller grid cells. The latter group are, almost be definition, microsimulation models – though as demonstrated by SimDELTA not all microsimulation models of household location work at the grid cell level. We consider some of the implications of microsimulation further below.

The need to disaggregate households, primarily on socio-economic criteria, is a common feature of the models considered. Within the aggregate models, the number of household categories varies widely, from eight in standard DRAM applications to over a hundred socio-economic/composition/employment status combinations in some DELTA applications. Microsimulation models which simulate the individual persons within households, as well as the household collectively, can to some extent avoid the need to define household categories in their actual location processes, though typically it is still necessary for model coefficients to relate to pre-determined categories.

In most models the total number of households to allocate across zones is exogenously prepared and specified as an input – though in forecasting applications this number is by definition itself a forecast and must come from some other form of model. In the DELTA case there is an intermediate stage in that household changes (formations, changes in composition and dissolutions) are calculated within the model, but using a household change model calibrated to match the results of more conventional demographic models at the study–total level. In Oregon2 and SimDELTA, the formation of households to locate is fully incorporated into the overall system and is driven primarily by the microsimulation of demographic changes (births, deaths, couple formation and dissolution) affecting the modelled population over time. In contrast, in UrbanSim the new households to be located by the microsimulation process in each year are synthesized to match aggregate demographic inputs.

2.1 Factors Influencing Residential Choice

Representation of the availability of housing is obviously central to the representation of residential choice in the most-developed countries where the vast majority of moves are into previously-owned dwellings or dwellings built speculatively by developers. Different factors would apply in economies where significant numbers of households have to create their own “informal” housing. DRAM is an exception amongst the models reported here in that its main supply variable is land rather than housing. Within the other models, the treatment of housing supply varies very considerably, from a single quantity of housing floor space in the DELTA applications through numbers of dwellings by type in many of the models. Tenure appears in some (e.g. TILT) but not all. The microsimulation models (Oregon2, UrbanSim and SimDELTA) inherently have the potential to consider significantly more detail, such as the number of rooms in an individual simulated dwelling, which is a key variable in SimDELTA. Whilst clearly more detail about housing supply should improve the residential location model itself, the addition of such detail implies the need to supply that detail (by zone and for each modelled year) in the forecasting process; as the DRAM paper points out, this can create issues both of practicality and of accuracy; this leads to the development of other sub-models to update all the other variables used (as for example in DELTA and MUSSA).

The influence of transport is present in each model in terms of travel to work (Calgary-Edmonton, DRAM, Oxford, TILT, UrbanSim, SimDELTA), travel to shopping (Calgary-Edmonton, Oxford), travel to school (Calgary-Edmonton, TILT), and less directly in accessibility measures (DELTA, SimDELTA).

Many other variables are considered to have an impact on residential location choice like air quality, street in front of dwelling (Calgary-Edmonton) quality of schools, noise (Oxford), room density (TILT), or zonal quality (DELTA), and neighbourhood characteristics (Albatross, Ramblas). Where fewer variables are used there is a tendency to make more use of the previously located numbers of households by category to imply values for influences which are not made explicit, either by an incremental formulation of the model (DELTA) which assumes that unmodelled variables remain constant, through other comparable formulations (MUSSA) or through the explicit inclusion of previous numbers of households (DRAM).

2.2 The Treatment of Dynamics

In the cases of the Edmonton-Calgary and Oxford models, they represent a state at a base year and therefore they are strictly cross-sectional. All of the other models involve at least some time-lagged terms. DRAM allocates population according to zone attractiveness calculated through period t values and based on zone-to-zone travel costs and the distribution of employment for period $t + 1$. In the case of DELTA, the time-lagged terms are combined with indirect use of a rent term, reflecting the balance of supply and demand in each forecast year, which therefore takes account of the changing demand for the stock of housing.

2.3 Issues of Interdependence

Most of the models calculate the probabilities that households will locate in each zone conditional on the household worker being employed in a particular zone; there is therefore a residential location process for each household type and for each work zone. In these cases, the implications for residential location of households having more than one worker are unclear. In the case of DRAM and the model for Oxfordshire, the location of jobs is itself endogenous to the model. The main exception to the general pattern is DELTA, which at the local level works on the basis that accessibility to work is just one of the characteristics influencing household location, and has a separate area-level migration process which draws workers towards areas with better work opportunities.

2.4 Representation of Planning Policies or Zoning Controls

In general, planning policies and zoning controls act as controls both on the processes of development and on the occupation of the resulting buildings. The processes of development and the physical supply of buildings for housing are essentially outside the scope of this book – authors were asked to focus on the processes by which households are located within a given housing stock. "Planning" policies in the conventional sense do not generally seek to control which households live in which dwellings. However, most Western societies have a range of "housing" or "social" policies which control the use of some housing and influence the use of the stock in general. These include, in particular, the provision of "social housing" which provides subsidised accommodation for households meeting certain criteria, and tax regimes which influence household preferences (sometimes by omission, for example by taxing capital gains on most forms of investment but not on owner-occupied housing). The ability to represent these kinds of policies requires first of all a distinction of housing by tenure, at least so as to define whether households are renting or buying their dwelling, and a substantial disaggregation of households by income/employment and age characteristics. Aggregate models such as DRAM or DELTA are consequently very limited in what they can do to represent such policies explicitly (though some of their consequences can be introduced implicitly, for example through the use of constraints as in the DRAM case. More fully dynamic models such as SimDELTA have the greatest scope in this respect, because the process of maintaining a household "history" over time allows for consideration of variables such as the outstanding value of the mortgage (housing loan) which an owner– occupier household owes on its dwelling.

Another example is represented by Albratoss and Ramblas where both models are primarily activity-based models of transport demand, and not integrated land use – transport models. Their prime goal is to predict activity–travel patterns and associated traffic flows. The distribution of residential land use, in terms of households and persons, is exogenously given. Based on the available data sources, a set of tools has however been developed to create synthetic populations that serve as input to the models.

2.5 Closing Observations

Our final point is to stress the variety of ways in which residential location modelling (and urban modelling more generally) is advancing. It includes both continuing refinements to model packages which have a long history, and wholly new developments, and demonstrates the very different ways in which the subject is being addressed.

One common misconception amongst some other groups of modellers is that LUTI modelling has advanced over the past three decades only by increasing disaggregation of the models originally developed in the "first generation" of operational models. The material presented in this volume, though selective and dealing only with one aspect of LUTI modelling, is more than sufficient to disprove this belief.

Another rather more sophisticated and more debateable view is that the historic trend in LUTI modelling is one of ever-increasing sophistication with large-scale microsimulation modelling as the ultimate approach. Whilst it is true that modelling a real 100% sample of households, persons and dwellings would clearly be an ultimate level of disaggregation, and that we can approximate this by modelling a comparable synthetic sample, the practical issues raised by microsimulation mean that it may not be the most desirable approach for application to forecasting in policy- and decision-making contexts. The points made in Putman's DRAM paper about the usability of models being a major determinant of their practical value remain highly valid. The most sophisticated models may therefore make more indirect contributions to the growing use of LUTI models in planning and government.

As editors we have been privileged to spend some time reading in detail about some of the work being done by our colleagues around the globe, and have been impressed yet again by the level of intellectual and practical effort being devoted to this form of modelling. This is only a partial picture; some of those invited to contribute to this particular volume were unable to participate, and we know that there are many others working in this area or on closely related models whom we were simply not able to consider. Like any such book in this age, this one will inevitably be out of date even before it appears in print; we would very seriously urge readers to seek updated information before relying on this book as describing either the state of the art in general or the state of any one model in particular. Despite these limitations, however, we very much hope that the book will be of value as a snapshot of the range of activity in this field, and that it will encourage others to join us in working on this perennially fascinating topic.

Acknowledgements The authors of this short discussion are very grateful to all the other authors represented in this volume for their input and for their patience. The present authors are wholly to blame for any errors of fact or opinion.

About the Editors

Francesca Pagliara is Assistant Professor at the Department of Transportation Engineering of the University of Naples Federico II.

John Preston is Professor at the Transportation Research Group, School of Civil Engineering and the Environment, University of Southampton.

David Simmonds is a Director of David Simmonds Consultancy Ltd, Cambridge, England.