# 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

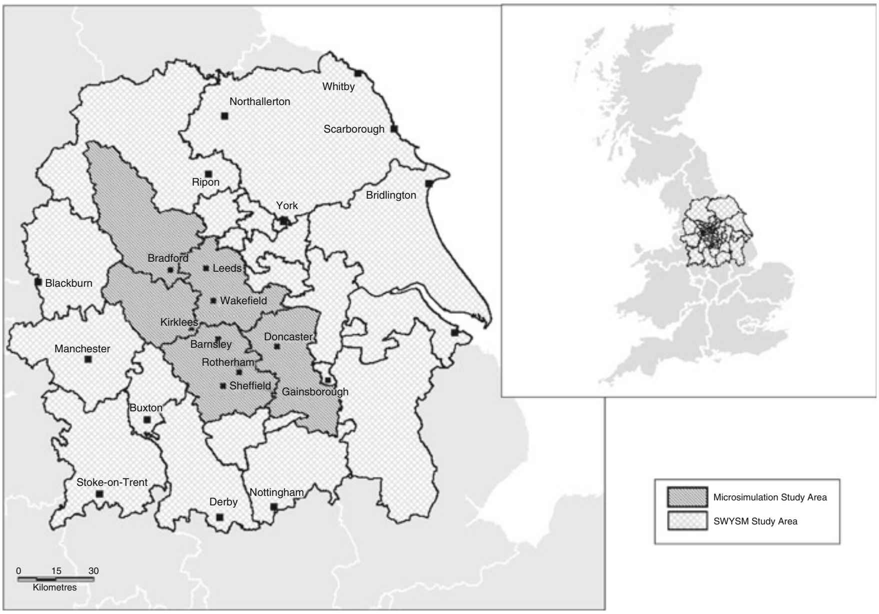


**Fig. 1** SimDELTA study area

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

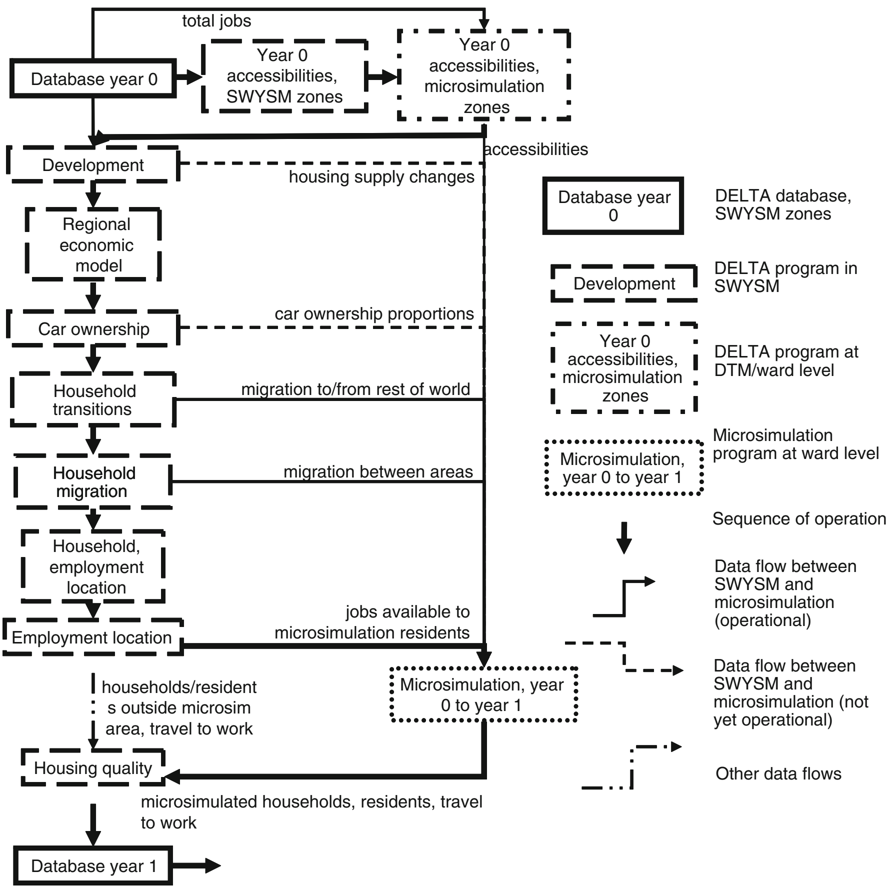


**Fig. 2** SimDELTA: linkages between aggregate and microsimulation components in 1 year Note that only the data flows affected by the addition of the microsimulation process are shown; there are numerous other flows between DELTA components and databases

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.

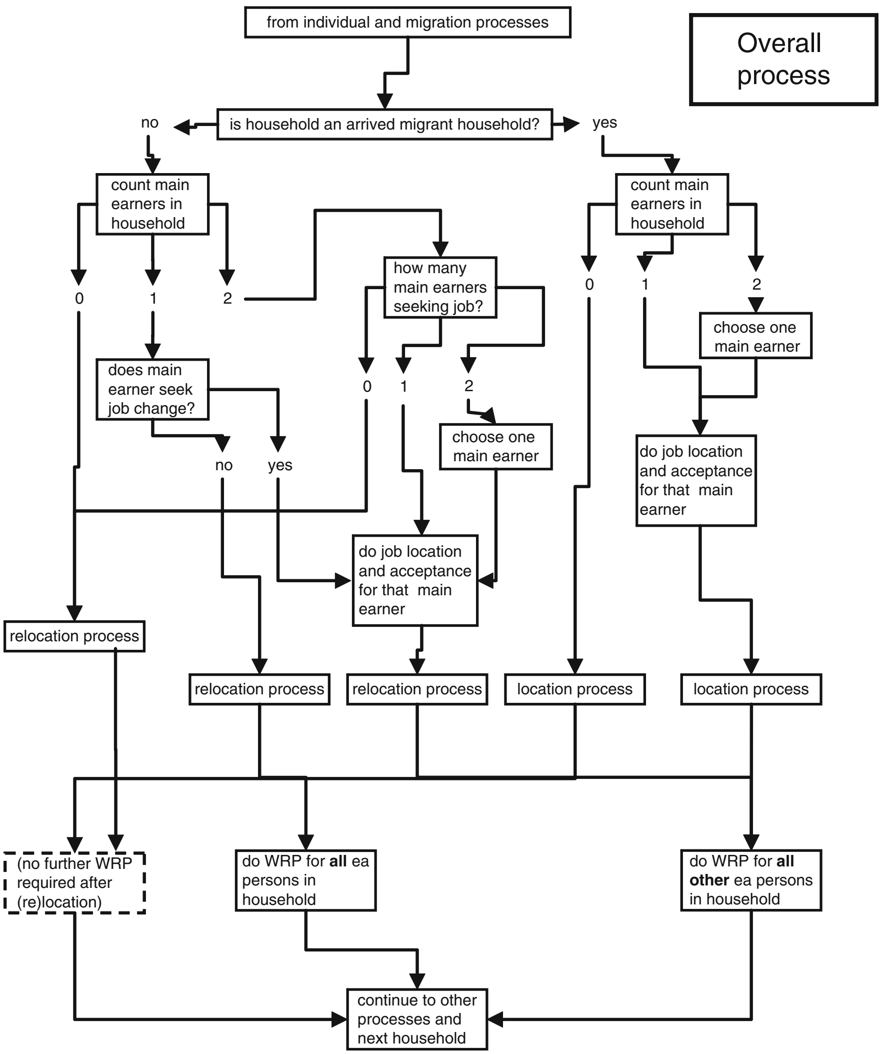


**Fig. 3** Overall process of residential/work choices (WRP – work related processes)

- 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