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 $m_{i_1 \dots i_m}$and marginal counts $C_{1(i_1)}, C_{2(i_2)}, \dots, C_{m(i_m)}$ 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, $m_{i_1 \dots i_m}$ 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 $m_{i_1 \dots i_m}$given marginal count $C_{j(i_j)}$ is according :

$$m'_{i_1 \dots . i_m} = m_{i_1 \dots . i_m} \frac{C_{j(i_j)}}{\sum_{i_1 \dots i_{j-1}, i_{j+1} \dots i_m} m_{i_1 \dots . i_m}}. \tag{1}$$

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 wV 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