# 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, 165
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.

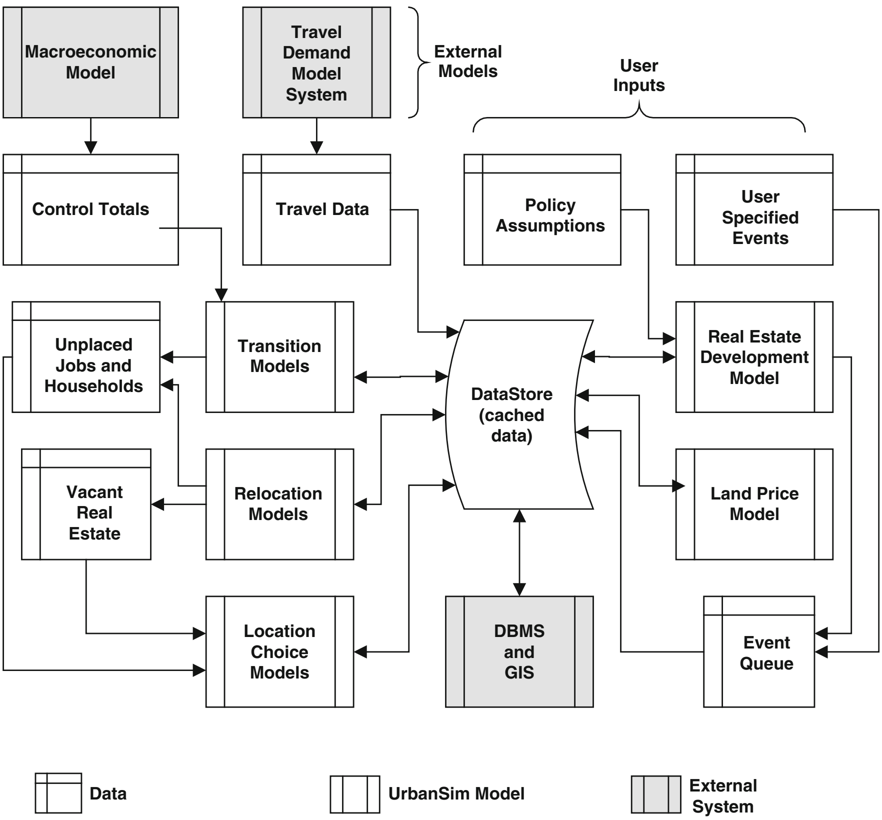


**Fig. 1** UrbanSim model system

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

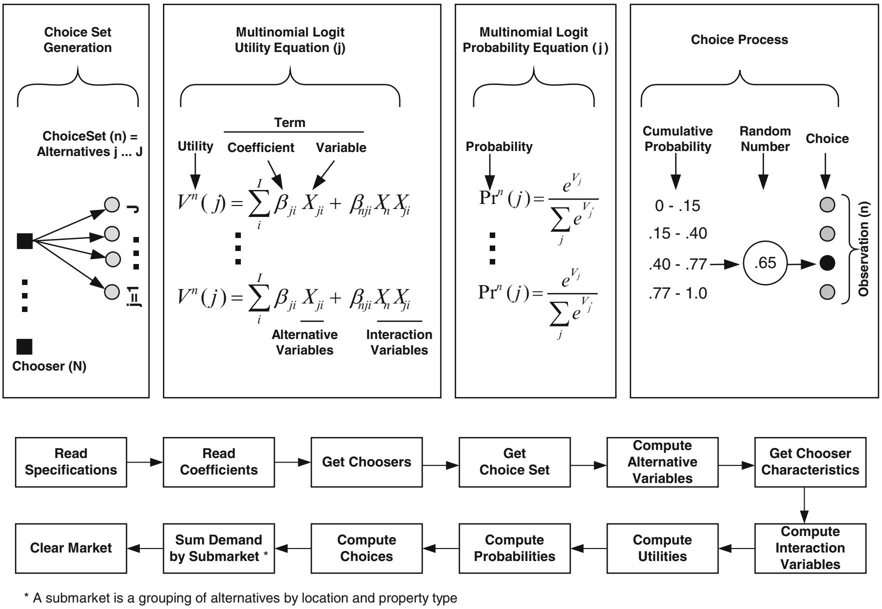


**Fig. 2** Computational process for household location choice model

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.

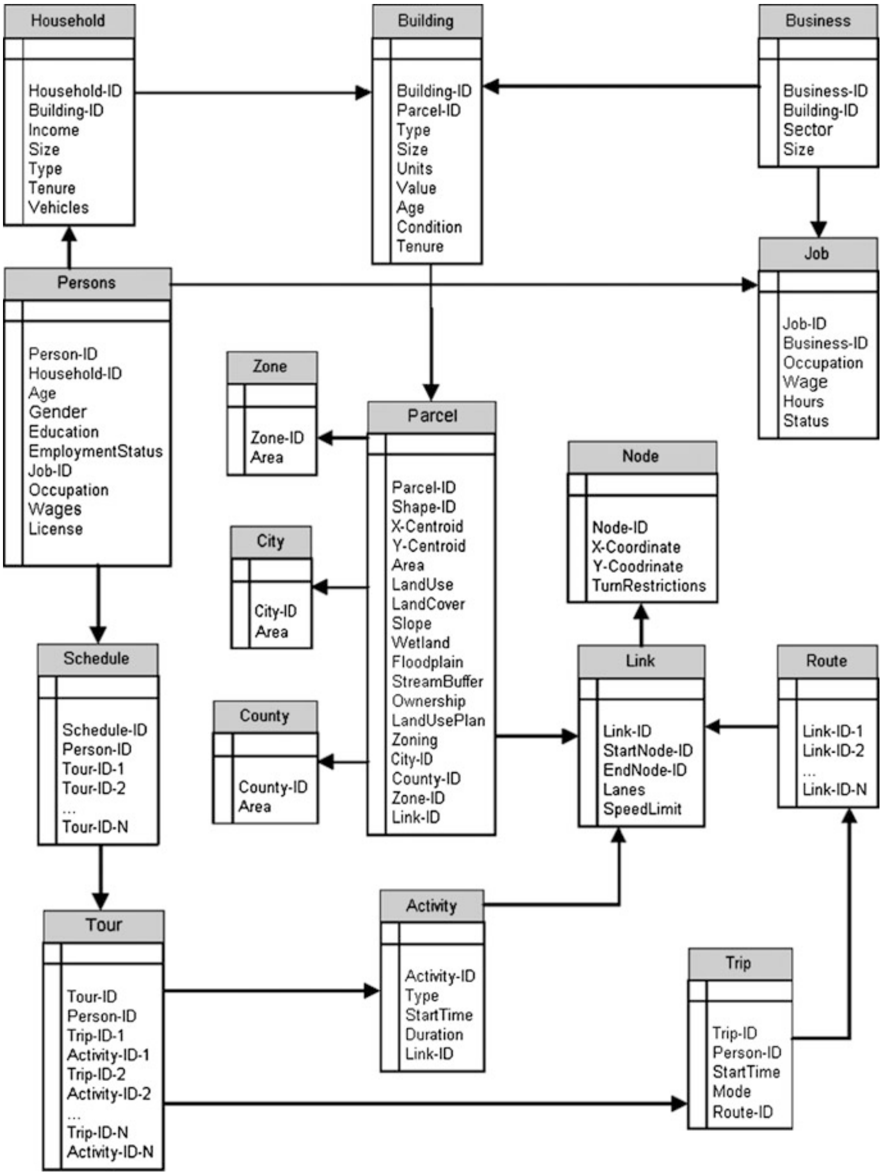


**Fig. 3** Entity–relationship diagram for an integrated parcel – activity based travel model system

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