Simulating household budgets for housing and transport

Rolf MOECKEL, Dr.-Ing.
Parsons Brinckerhoff
6100 NE Uptown Blvd., Suite 700
Albuquerque, NM 87110
United States
Tel: +1 (505) 878 6553
E-mail: moeckel@pbworld.com

Abstract
Today’s hot topics in urban planning, such as smart growth, high energy prices or congestion pricing, require integrated land-use/transportation models to analyze the impact of such scenarios. The majority of land-use models available today are either not integrated with transportation models or require a large amount of data and are challenging to calibrate. SILO (Simple Integrated Land-Use Orchestrator) is a robust yet simple land-use model that can be fully integrated with a transportation model. It is designed to be built with incomplete data and limited time and monetary resources. SILO has been designed as a microsimulation, simulating each household individually. This helps representing the interaction between single agents, and it further allows integrating SILO with both aggregate (4-step) and disaggregate (activity-based) travel demand models.

The pilot application has been implemented for the Minneapolis/St. Paul Metropolitan Area. The model currently simulates households (including moves, birth, death, marriage, divorce and leaving the parental household) and the dwelling real-estate market (including construction, renovation, deterioration, demolition and change in rent). It is planned to extend the model to simulate employment and the non-residential real-estate market as well.

In anticipation of high energy prices, the model is designed to ensure that households do not exceed their monetary budgets. If commuting under high energy prices exceeds the travel budget, the household will move either to a location closer to work or the household will move to a less expensive dwelling and rearrange the share between housing and travel budgets. This method ensures that households do not exceed their budgets, even under severe energy price increases.

Keywords: Land-use modeling, household budgets, integrated land-use/transport modeling
1. INTRODUCTION

The rising demand for analyzing policy questions that address environmental aspects of planning has fostered the interest in integrated land-use transportation models. Analyzing the effects of transit-oriented developments (TOD), mixed land-use, fuel tax, fees on vehicle-miles traveled (VMT fees) or higher energy prices require the integration of land-use with transportation models to capture the entire feedback cycle as described by Wegener (1994) and shown in Figure 1. While the transport system provides travel times and accessibilities that influence location decision of households and businesses, the location of land uses defines the origins and destinations of trips and, therefore, is the driver for the use of the transportation system. An intriguing example showing the benefit of integrating land use with transportation modeling has been published by Conder and Lawton (2002). They describe that a Delphi panel had assumed large growth north of the Columbia River in the Portland, Oregon Metropolitan Area. Using an integrated land-use/transportation model, in contrast, showed that the additional residential development north of the Columbia River would lead to a serious congestion of the bridges across the river, making living north of the river much less attractive than assumed by the Delphi Panel.

Short of integrating land-use with transportation models, the effects of many policies are likely to be misrepresented. If, for example, the impact of TOD on travel demand is analyzed by simply shifting households from auto-dependent locations into TOD housing, the travel demand model would recognize the higher transit accessibility and reduce auto travel demand. By contrast, using a land-use model to simulate which households are likely to move into TOD dwellings may show that there are specific households that move into such a neighborhood, likely having a different travel behavior than the traditional suburbanite.

While travel demand models have undergone a remarkable development in the last two decades from simple three-step models to four- or five-step models or activity-based models, land-use models remained a niche for rather academic applications. In North America, very few implementations of integrated land-use transportation models reached a level of maturity that allowed running a larger number of scenarios.

One reason why land-use models are not as widely applied as transportation models is the limited availability of land use models that can be integrated with transport models yet do not require a lot of time and budget to be implemented. Excellent reviews of land use models have been written by Wegener (2004), Hunt, Kriger et al. (2005) and the U.S. Environmental Protection Agency (2000). The majority of these models are sophisticated
academic exercises that have largely defined the state-of-the-art in land-use modeling, however, only few of them have been applied to more than one demonstration study area, and many of them lack proofing the ability to run a larger number of scenarios for policy analysis.

In North America, it appears as if the practice of land use modeling is divided into two very different approaches. On the one hand, there are very simple sketch-planning tools that allow visualizing and summarizing land development scenarios developed by expert panels. Examples are the land-use models Smart Growth Index (Criterion Planners, 2002), I-Place’s (Orton Family Foundation and Placeways, 2010) or What-If (Klosterman, 1999). Though these tools are helpful for finding consensus among different stakeholders, they lack the integration with a travel demand model. On the other hand, there are very comprehensive and complex land-use models, such as PECAS (Hunt and Abraham, 2003), TRANUS (de la Barra, 1989) or UrbanSim (Waddell et al., 2003). Though these models are successful at simulating land-use changes with a high degree of detail, they tend to be data hungry, have long run times, and lock behavior from the past into the model through rigorous econometric parameter estimation and calibration.

There is a lack of land-use models that are fully integrated with a travel demand model but simple enough to be implemented with a small budget of time and/or funding. Data requirements need to be reasonable and in line with common data availability for most regions.

2. CONCEPT

To allow integrating land-use modeling with travel demand models in a limited data environment, a simple yet robust land use model has been developed. It is called SILO, which stands for Simple Integrated Land-Use Orchestrator. The primary purpose of SILO is to provide the land-use side for integrated simulation models. Highly congested highways or bridges may prevent the intended growth of certain areas. Mixed-use developments may attract different residents with a different travel behavior. SILO is designed to implement these interactions with a travel demand model.

SILO is built to simulate a large number of policy scenarios in short run times. Foremost, changes in zoning can be implemented by restricting certain developments in selected areas. The model is designed to be sensitive to changes in transportation costs. If prices for gasoline or tolls are raised significantly, the model shows that households choose dwellings that are closer to their daily activities and less car-dependent. Overrides can be implemented to add or remove development of dwellings and non-residential floorspace as exogenous input. Overrides are used to simulate the effects of changes in land use that are assumed to be certain. Sometimes, these are called historic events, such as the immigration or outmigration of a major employer. Overrides also can be used to analyze the effects of a planned job/housing balance.

The simulation results can be summarized by any characteristic included in the model design. Besides distribution of households and population, the user may analyze how
different socio-economic groups are affected by certain policies. If the user is interested in assessing the demand for schools or elderly housing, the model allows summarizing the data by the corresponding attribute. While the model will not be able to predict demand of single facilities, which would be a violation of microsimulation theory, the total demand in the entire study area can be compared to the available capacities.

The model works incrementally instead of reinventing the distribution of population and employment from scratch every simulation period. The latter is common in projects where a synthetic microscopic population is needed by a travel demand model in each simulation period. Commonly, this is implemented by using an equilibrium-based approach that distributes population and employment to maximize the location utility (Müller and Axhausen, 2011). This process tends to overestimate the willingness of households and firms to relocate. In reality, both households and businesses are hesitant to move and only relocate if an alternative location provides a large gain in utility over the current location. When implementing an incremental change, in contrast, the base year population is only slightly adjusted every simulation period, similarly to changes occurring in reality. By only adjusting the part of the population that actually moves, a certain disequilibrium is kept on purpose to closer resemble real-world behavior.

The model is developed as a discrete-choice model based on theory developed by Domencich and McFadden (1975). The discrete-choice theory is implemented by logit models that represent decisions made under uncertainty. The logit model does not assume full market transparency, but choices are simulated as decisions made under uncertainty with limited time and money and a personal bias due to perceptions, prejudices and habits. Simulated decisions rarely maximize the utility but rather satisfy the needs. For land-use modeling, this approach often is considered to be closer to real-world behavior than equilibrium-driven approaches.

The land-use model consists of four modules: households, jobs, dwellings and non-residential floorspace (Figure 2). The former two have been implemented, and the latter two are only designed so far. The model can be applied as a fully integrated land-use/transport/environment (LTE) model. While simulated population and employment are used as generators and attractors for the travel demand model, travel times are converted into accessibilities and fed back into the land use model. Both land use and transport generate gaseous emissions and noise that can be estimated and be fed back into the land-use model as another location factor: environmental quality.

The model is developed as a microsimulation. This allows integrating the model with both aggregate 4-step travel models and activity- or tour-based models. Microsimulation land-use models have proven to represent the interaction between single actors, they can be executed within small run times and are no more difficult to implement than aggregate models. A potential drawback of a microscopic approach is the dependence on random number generators (Wegener, 2009). Stochastic variability may be significant if working in smaller areas or analyzing single zones. To address this issue the model is designed to complete in a short runtime, allowing the completion of a large number of runs with identical model settings. This way, the variability due to stochastic variation can be analyzed. If the variability is small at the geographic level of analysis, it is acceptable to use an average model run. Gregor (2006) proposes to keep every run of the microscopic model LUSDR as a possible scenario outcome. Donnelly (2009) ran a
microscopic commercial vehicle model 60 times keeping track of the random number seed of each model run. As the final model output a run was chosen that closely resembles the average of all 60 model runs. Both Gregor's and Donnelly's approach help selecting a model run that is representative. To avoid a stochastic bias in SILO, two paradigms are followed. First, several model runs with the same settings but different random number seeds are completed and the variability is analyzed. Secondly, no geographically detailed data are extracted for scenario evaluation. Scenarios have to be analyzed at a larger geographic scale to avoid reaching conclusions that could be based on stochastic variation.

Figure 2: Concept of the SILO Model

SILO is implemented using zones as its spatial resolution. In the last decade, several land-use models were implemented at the parcel level. Though this is an admirable effort, a true benefit of dealing with parcels instead of zones has – to the knowledge of the author – not been proven yet. Working at the parcel level is part of the reason why existing integrated land-use model tend to be difficult to implement. To reduce the data demand, SILO is implemented at the zonal level. If future applications require a higher spatial resolution, raster cells might provide all detail necessary while limiting the requirements for data collection.

Another key component of SILO is the explicit representation of monetary budgets. Budgets are considered to be a major constraint when organizing activities. As Wegener points out in his space-time theory, everybody has 24 hours a day and everybody has to get by with her or his income in the long-run (Spiekermann and Wegener, 2009). Zahavi (1974) proved statistically that the time budget for trips is rather constant and may change only slowly over time. Therefore, time and monetary budgets are tracked for every household. If an employer moves from a central location to the suburbs, and therefore, increases commute times for many of its employees, some employees will relocate to reduce the commuting burden. If energy prices rise and transportation will become more expansive, keeping track of travel costs helps SILO to trigger a move if costs for transportation, housing and other costs exceed the household income.
A graphical user interface (GUI) has been developed to control the model. The GUI selects the years to simulate, the components to include, a scenario name and the files that contain the input data.

**Major design features of SILO**

- Emphasize on small runtime. 30 simulation periods are run in under two hours, allowing to test a large number of scenarios
- Limited data requirements. More emphasis is put on reasonable model responsiveness than elaborated data collection and calibration. Most model parameters are assumed heuristically and may change over time.
- Explicit representation of household budgets. Every household has to get along with its budget. A household may not chose a dwelling that would in combination with the household’s transportation costs exceed the household budget.

### 3. SIMULATION MODULES

Households are simulated in two submodules called demography and moves. Demography covers the most important demographic events, including birth, aging, marriage, divorce, children leaving the parental household and death. Events are simulated in random order to avoid path-dependency. Markov models are used to simulate these transitions with age-, gender- and household-based transition probabilities. The moves module covers the relocation of households in two steps (Moeckel and Osterhage, 2003: 108-114). A logit models simulates the decision whether
to look for a new home or not, which is based on the perceived improvement from moving.

\[ p_{move} = \frac{c}{1 + e^{c(\beta (1 - \Delta u))}} \]  

(1)

where \( p_{move} = \) Probability to move  
\( c = \) constant  
\( \beta = \) parameter  
\( \Delta u = \) expected average improvement in housing utility

To calculate the expected average improvement, the utility of all available vacant dwellings is compared with the utility of the household's current dwelling. The numerator of equation 2 has the average utility improvement if the household decided to move, while the denominator has the change of utility if the household stays in the current apartment (which is equal to zero):

\[ \Delta u = \frac{\sum_j \exp(\mu(u_j - u_i))}{\sum_j \exp(\mu \cdot 0)} \]  

(2)

where \( j = \) Available vacant dwelling  
\( \mu = \) parameter  
\( u_i = \) Utility of dwelling \( i \)

If the household decides to move, another logit model selects a new neighborhood. Selecting a neighborhood before selecting a dwelling is an important step when using the logit model. If all vacant dwelling were compared in a single logit model, a base assumption of the logit model would be violated (Domencich and McFadden, 1975). The logit model was developed to compare a limited number of options, similarly to the ability of a household to compare a few dwellings only at one point of time. Therefore, a new housing location is chosen in two steps, first the neighborhood and following the dwelling is selected. The equation for both steps is similar, with the exception that first the utility of a neighborhood is used, and secondly, the utility of a dwelling is analyzed. The following equation expresses both steps, where \( i \) and \( j \) either represent a neighborhood (first step) or a dwelling (second step).

\[ p_{i,j} = \frac{A_j \cdot u_j \cdot \exp(\gamma(u_j - u_i))}{\sum_i A_i \cdot u_i \cdot \exp(\gamma(u_j - u_i))} \]  

(3)

where \( p_{ij} = \) Probability to move from \( i \) to \( j \)  
\( A_j = \) Number of available dwellings in \( j \)  
\( u_j = \) Utility of neighborhood (first step) or dwelling (second step) \( j \)  
\( \gamma = \) parameter

The utility of a neighborhood is comprised of the median price of dwellings, accessibilities and environmental quality:
\[ u_n = \alpha_1 \cdot \bar{p}_n + \alpha_2 \cdot acc_n + \alpha_3 \cdot env_n \]  

where  
- \( u_n \) = Utility of neighborhood \( n \)  
- \( \alpha_1, \alpha_2, \alpha_3 \) = Parameters, adding up to unity  
- \( \bar{p}_n \) = Median price of dwellings in neighborhood \( n \)  
- \( acc_n \) = Accessibility of neighborhood \( n \)  
- \( env_n \) = Environmental quality of neighborhood \( n \)

Dwelling utilities are based on price, size, and quality of the dwelling as well as a binary budget indicator:

\[ u_i = p_i^{\beta_1} \cdot s_i^{\beta_2} \cdot q_i^{\beta_3} \cdot bl \]

where  
- \( u_i \) = Utility of dwelling \( i \)  
- \( \beta_1, \beta_2, \beta_3 \) = Parameters, adding up to unity  
- \( p_i \) = Utility of price of dwelling \( i \)  
- \( s_i \) = Utility of size of dwelling \( i \)  
- \( q_i \) = Utility of quality of dwelling \( i \)  
- \( bl \) = budget limit: 1 if costs for housing + transportation + other needs <= household income, 0 if costs exceed household income

Note that the utilities were added at the neighborhood level in equation 4 (weighted addition) and multiplied at the dwelling level in equation 5 (Cobb-Douglas Function). The addition at the neighborhood level expresses that a poor utility of accessibility may be made up by a high accessibility of price, which represents the trading off between a more expensive very accessible location and a less expensive and less accessible neighborhood. In contrast, utilities at the dwelling level are multiplied to express that all elements need to be fulfilled to a certain degree. For example, if the utility of the price is zero (i.e. it is too expensive for this household), or the utility of the size is zero (i.e. it is too small for this household), the utility of the entire dwelling becomes zero, and therefore, is dismissed by this household.

The term \( bl \) in equation 5 ensure that the costs of a dwelling plus transportation costs do not exceed the household income. Transportation costs are calculated based on the location of that dwelling and the current work locations of all household members. Multiplying the round-trip distance with current transportation costs (which is an exogenous input) provides transportation costs for work trips. The amount is increased by a fixed share (currently set to 20 percent) to account for further mandatory trips, such as grocery shopping or doctor visits. If these transportation costs plus housing costs plus other costs of living exceed the household income, the factor \( bl \) is set to 0 and the entire utility of this dwelling becomes 0. In an extreme scenario with very high transportation costs, it may happen that not all households find dwellings they can afford. The model reports how many households did not find affordable housing. Developers will focus investment on less expensive housing in locations that allow lower transportation spending. Such a scenario will also severely reduce the number of children leaving the parental household. In subsequent scenarios, landlords may reduce the costs for housing in inaccessible or car-dependent neighborhoods as a reaction to increased transportation costs.
The utilities are evaluated based on household type preferences. Currently, 20 different household types and their housing preferences are distinguished. In line with the intent of SILO to reduce the effort of implementing the model, housing preferences are not estimated econometrically but rather assumed heuristically. In other words, the pilot application applies universal urban theories rather than estimating coefficients, which are by definition time- and location-dependent. Heuristically assumed parameters may change over time, and thus do not lock behavior of the past into future simulated years.

Household moves are decisions of special interest in urban modeling, and therefore, simulated by logit models. Demographic changes, in contrast, are simulated by Markov models. Markov models apply transition probabilities without further analysis of the reasoning for the decision. As commonly no scenarios are tested that aim at changing demographics, Markov models are an appropriate simplification for demographic transitions.

Dwellings are defined by type, size, price and quality. Dwelling types distinguished are single-family homes, multi-family homes and mobile homes. Size is defined by number of bedrooms (0, 1, 2, 3, 4 or more). Depending on the simulated demand, investors decide to develop new dwellings. The location choice for the new building is simulated by logit models. Developers use similar location choice criteria as households, as they anticipate the preferred location of future renters or buyers. Markov models are used to simulate transitions of dwellings, including renovation, deterioration or demolition. Land-use regulations may restrict the development of housing in certain zones. In addition, the model allows overrides. If the user knows about specific housing developments, new dwellings can be added exogenously to single zones and existing dwellings may be selected exogenously for demolition.

The simulation of employment is not implemented but designed. Employment will be represented in a simplified form by simulating their jobs instead of firms. The number of industry types can be set by the user, but commonly 4 to 10 types will be distinguished. Firmographic changes, i.e. the demography of firms, will be simulated by Markov models and include job creation and job termination. Jobs may relocate if their location utility improves at a different premise significantly. Location decisions will be simulated by logit models, evaluating the quality of non-residential floorspace, the neighborhood and the price. The quantity of floorspace used by one employee needs to be given exogenously for each industry and may change over time.

Non-residential floorspace will be simulated similarly to dwellings. This floorspace is distinguished by type, quality and price. Developers decide based on the floorspace demand if they develop additional floorspace. The location choice for new floorspace will be simulated by logit models. Land-use regulations may restrict the development in certain areas. Markov models will be used to simulate renovation, deterioration and demolition of non-residential floorspace.

The number of moves of households and jobs from outside into the study area and from the study area to the rest of the world is not simulated endogenously, as it depends on utilities of other regions outside the study area. Thus, external moves are given as an exogenous input by household type and employment industry.
The base year prices of dwellings and non-residential floorspace are derived from land price data. Depending on simulated demand and supply, prices are updated at the end of each simulation period. Price increases are assumed to happen instantaneously, whereas decreases happen only if broad vacancy persists. This reflects common behavior of landlords who adjust prices upward more quickly than reducing prices in times of low demand. Prices are updated by a regression model considering recent growth, current supply and current demand for each dwelling type.

4. STUDY AREA

A prototype of SILO is implemented for the Minneapolis/St. Paul Metropolitan Area in Minnesota. Figure 4 shows zonal population density of the 1,201 zones of the study area. The region has a population of approximately 2.4 Million living in about 1 Million households.

This study area was chosen for a couple of reasons. Foremost, the staff of the MetCouncil Metropolitan Planning Organization is providing great support in developing this model in terms of sharing input data and discussing the model design. Furthermore, the metropolitan area has the right size for a prototype land-use model application. While it is large enough to go through major urban development phases, it is not a mega-region that requires a disproportionally large amount of data collection. Finally, and maybe most importantly, there is another land-use model under development for the same area. Citilabs is implementing Cube Land, which is a new land-use model based on the MUSSA (Martínez, 1996, Martínez and Donoso, 2007) model. Having Cube Land implemented in the same area will offer the great benefit of comparing two model results in a meta-analysis. As shown by Wegener et al. (1991), such meta-analysis helps understanding the impact of model design on model outcome. By no means is SILO meant to compete against Cube Land. Much rather having a second land-use model in place, will largely benefit the development of SILO, and possibly of Cube Land as well.
5. ESTIMATION, CALIBRATION AND VALIDATION

The first prototype of SILO for Minneapolis/St. Paul is built with readily available data. In some cases, local data were available (such as number of households, number of dwellings, or estimates of land prices). In other cases, nationwide data were used instead (such as birth rates or moving rates). This prototype application clearly focuses on model development rather than ambitious data collection. To a large extent, the estimation and calibration of model parameters is done heuristically, meaning that more focus is spent on applying reasonable parameters rather than estimating those econometrically.

A series of sensitivity tests is assumed to be more important than exhaustive calibration. Sensitivity testing shows that the model reacts reasonably to changes in population and employment growth rates, urban growth boundaries and changes in the transportation system.

Finally, the validation shows that the model is able to approximate real-world development. The base year of the model is set to 2000 to allow back-casting. The model is run from 2000 to 2008 and model output is compared to the population distribution in 2008 American Community Survey (ACS) data. Successful back-casting increases confidence that the model is able to predict future development.

6. DATA REQUIREMENTS

It is common to start a land-use paper with data requirements. In this case, however, data requirements are purposefully moved to the end of the paper, as model design and reasonable representation of behavior are valued higher than a detailed data collection. The model is developed with the least amount of data necessary, and additional data are collected as needed, rather than spending a lot of time on data collection at the outset. Missing data need to be replaced with plausible assumptions. Extensive sensitivity testing allows assessing how severe the lack of detailed data might be. This helps focusing the data collection effort on those pieces that are most influential in the modeling stream, rather than spending a lot of effort on data collection that might have little impact on model output.

A common misunderstanding in practice is that land-use models would require a tremendous amount of input data. This approach demonstrates the opposite. At the beginning, a synthetic population has to be generated. Sample enumeration has been applied to expand PUMS (Public Use Microdata Sample) data to a full synthetic population of the study area. PUMS data are microscopic and published by the U.S. Census Bureau, the most recent data represent the year 2000. Households are represented, among others, by household size and dwelling ID. Each person has attributes, such as age, gender, relationship to head of household, education, occupation, workplace location or income. The corresponding dwelling of each household is
described, among others, by a PUMA (Public Use Microdata Area) Zone, a weight, the dwelling type, number of bedrooms and rent or mortgage. The location needs to be disaggregated from one of the 20 PUMA zones covering the Minneapolis/St. Paul Metropolitan Area to its 1201 zones. MetCouncil provided the number of dwellings by type at the zonal level, which is used as a weight to disaggregate from PUMA zones to the corresponding model zones. In addition, vacant dwellings needed to be created to match the total number of dwellings by zone given exogenously. Single PUMS records within the right PUMA zone of the right dwelling type are selected randomly to fill up the real-estate housing stock with empty dwellings to match the control total.

Missing data pieces need to be synthesized. For example, the PUMS data does not provide information on auto-ownership. Autos are selected for each household based on location, income and household size. Finally, forecasts of population and employment growth for the entire study area are needed. As growth of the Minneapolis/St. Paul region is influenced by the growth of other regions in the United States and abroad, no effort is made in simulating growth in the Minneapolis/St. Paul region endogenously. Exogenous forecasts are used instead.

7. CONCLUSIONS

The concept of the model presented here relies on the application of urban theory, rather than the extensive collection of input data. This ensures that the behavior observed in the base year is not locked into the model using constants. Instead, parameters may change over time based on urban theory, such as a declining birth rate, more emphasis on school quality for housing location search, or growing auto-ownership. Furthermore, it allows developing and applying the model with smaller funding and in a shorter timeframe. The key question for every data element added is whether it helps improving model results. No data element is added for the mere sake of providing a more detailed simulation.

The ability to run a larger number of model runs is crucial for SILO. This helps performing the heuristic calibration, as many model runs are needed to fine-tune the model. In addition, a large number of scenarios helps selecting the preferred planning alternative. Finally, the average of multiple model runs is required to avoid stochastic variation of single model runs. Therefore, high priority is given to short run times. While SILO completes 30 simulation periods in less than two hours, most travel demand models need much more time for a single simulation period. In most applications, it won't be able to run the travel demand model more frequently than every 10 simulation periods. Though it is not idea to update the travel times every 10 years only, this is a necessary simplification to keep runtimes manageable. In an ideal world, an integrated land-use/transportation/environment model should not run longer than overnight.
8. REFERENCES


CRITERION PLANNERS 2002. SMART GROWTH INDEX. A Sketch Tool for Community Planning.


