Modeling Real Estate Demand

Webinar 6 of an 8-part TMIP Webinar series on land use forecasting methods.

Paul Waddell, 2011

Land Use Forecasting Webinar Series

1. The Evolving State of the Practice
2. Land Use Theory and Data
3. Scenario Planning and Visioning (I-PLACE3S)
4. Spatial Input-Output Frameworks (PECAS)
5. Dynamic Microsimulation (UrbanSim)
6. Modeling Real Estate Demand
7. Modeling Real Estate Supply
8. Scenario Planning and Visualization

2

Paul Waddell, 2011

Webinar 6: Modeling Real Estate Demand

1. How Much Detail Do We Need in Models: An Incremental Approach
2. Disaggregate Location Choice Modeling: A Discrete Choice Approach
3. Case Studies in Model Development

a) Housing Tenure and Location, Zone Level, San Francisco Bay Area Household Location Choices at Parcel/Building Level, Puget Sound Region
b) Modeling Employment Location Choices

Paul Waddell, 2011

Webinar 6: Modeling Real Estate Demand

1. How Much Detail Do We Need in Models: An Incremental Approach
2. Disaggregate Location Choice Modeling: A Discrete Choice Approach
3. Case Studies in Model Development

a) Housing Tenure and Location, Zone Level, San Francisco Bay Area Household Location Choices at Parcel/Building Level, Puget Sound Region

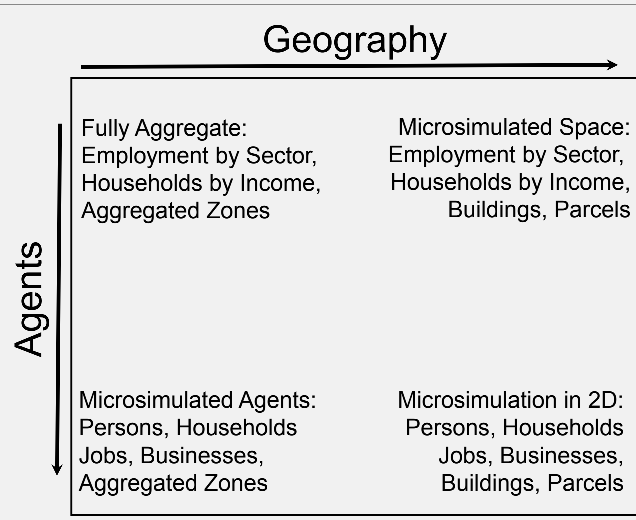
Paul Waddell, 2011

How Much Detail do we Need in Integrated Modeling?

- How much detail do we need in land use and transportation models in order to achieve models that are unbiased and allow us to assess relevant transportation and land use policies?
  - How small should zones be?
  - Should we use zones at all? What about using parcels, or small gridcells?
  - How much detail do we need in terms of population and employment?
  - Should we use aggregations of households and jobs, or microsimulate?
- Which details are important?
- What biases might our models have if we leave out details?
- What errors might we introduce if we use too much detail?

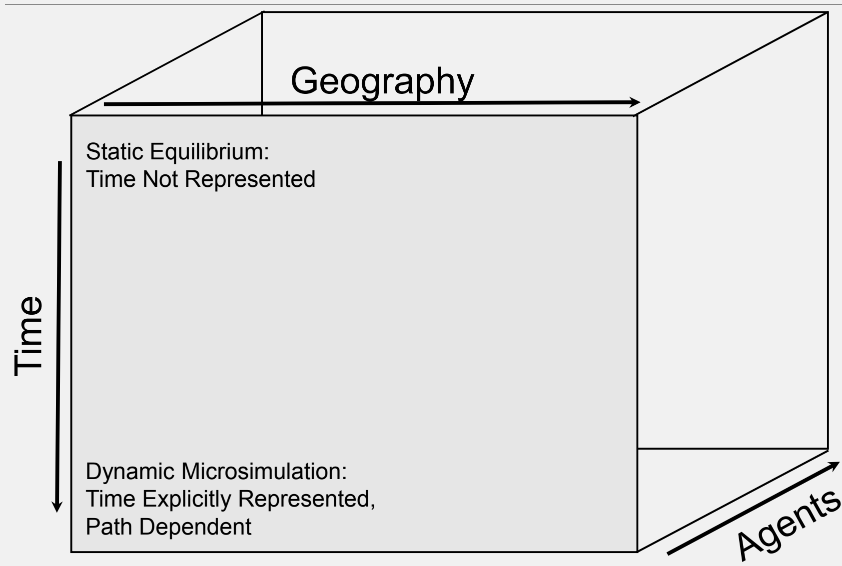
Paul Waddell, 2011

Detail in 2 Dimensions



Paul Waddell, 2011

Detail in 3 Dimensions



Paul Waddell, 2011

To Microsimulate, or Not to Microsimulate. That is the Question

- Traditional urban models have ALL been aggregate
- Economic models use a representative agent to motivate model
- Urban economics is based on the Monocentric model
  - Disaggregated at most by industry, households by high vs low income
  - Analytically based models are not tractable with much detail
- Later, more applied models began disaggregating:
  - DRAM/EMPAL: households by income quartile, employment by sector
  - TRANUS/MEPLAN/PECAS: use more sectors and household categories
- Microsimulation models are fundamentally different:
  - Originated with work of Guy Orcutt, initially for policy analyses like tax incidence
  - Represent individual agents (households, persons, jobs, businesses)
  - Maintain these as lists, and update them as the model progresses

To Microsimulate, or Not to Microsimulate. That is the Question

- Arguments used in favor of aggregating individual agents:
  - Data is more compact and easy to use (when small numbers of types used)
  - Leads to simpler models (depends on model logic: can be very complex / black box)
  - Models run faster (this is debatable and depends on complexity and implementation)
  - Less prediction error due to aggregation (this is an empirical question)
- Arguments used in favor of microsimulating individual agents:
  - Data are more natural to understand and work with since they represent real agents
  - Models can be more straightforward in their logic: agents make choices
  - Models can be designed to run at least as fast as aggregate models (especially aggregate models with many categories of agents and outcomes)
  - Avoids Ecological Fallacy: a common error in social science research, where individual behavior is inferred from aggregate data
  - Model parameters are less likely to be biased
- Field is moving strongly in the direction of microsimulation of agents

Time is of the Essence

- Urban models have generally ignored time
  - Time-abstract models use a notion of static equilibrium from economics
  - We assume that a city-region is in equilibrium, perturb the equilibrium, and observe a new static equilibrium
  - Attractive for theoretical analysis since it follows from theory and provides consistent answers every time
- Recent models have begun to represent time explicitly
  - Empirical observation suggests that markets may often be in sustained disequilibrium: subprime mortgage crisis and subsequent global recession
  - Increasingly common to represent annual time steps reflecting differing response times in real estate supply, intra-year changes in household location, business location, prices
  - Path dependence is a feature of this kind of modeling: changes today have implications for later choices (developers go bankrupt because of imperfect foresight)
- Field is moving strongly in the direction of explicit representation of time

Paul Waddell, 2011

Location, Location, Location

- Most common arguments in favor of using zones vs parcels are:
  - Less data requirements
  - Easier to develop
  - Easier to diagnose
- Most common arguments against using zones:
  - Modifiable Aerial Unit Problem (MAUP): model results depend heavily on configuration of zones; parameters sensitive to zone configuration
  - Ecological Fallacy: easy to fall into a classic mistake in social science research: inferring individual behavior from aggregate data
  - Walking scale is below the radar: may bias models with respect to intra-zonal trips, non-motorized trips, transit trips (with walk access); this is compounded by using travel networks that exclude local streets

Paul Waddell, 2011

Location, Location, Location

- Most common arguments in favor of using parcels vs zones are:
  - Data is becoming more readily available from tax assessors and commercial sources
  - Parcels are real: they are the unit of land that is owned, subdivided, and developed
  - Easier to interface local jurisdictions plans and zoning with regional plans
- Most common arguments against using parcels:
  - Messy data
  - Large data storage and processing
  - Difficult to standardize across jurisdictions
  - For larger regions, can take 2+ years to make parcel-level data useful for modeling using generally available data techniques

Paul Waddell, 2011

Recommendation: Start Simple, Then Add Details

- Experience has demonstrated that an incremental development path may be most productive:
  - Begin with a very simple zone model configuration
  - Begin using and assessing the model
  - Add detail to the model incrementally, targeting most pressing needs
- Example: Incremental development using a single platform - UrbanSim
  - Use microsimulation of agents and explicit representation of time, just vary geographic detail
  - Modular configuration and estimation of models
  - Zonal model configurations
  - Synthesizing parcel details
  - Parcel model configurations
  - Exploiting parcels and local streets

Paul Waddell, 2011

UrbanSim: Start From the Simplest Zonal Configuration

Household Location Models

Household Transition Model

Household Location Choice Model

Employment Location Models

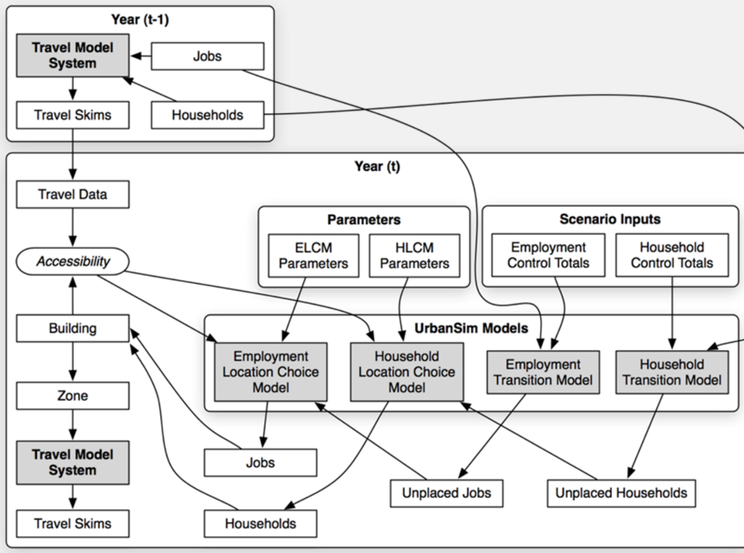
Employment Transition Model

Employment Location Choice Model

No representation of supply side of real estate market, or prices. No relocation of agents once placed. Becomes an 'incremental' model, allocating growth.

Paul Waddell, 2011

The Simple Zone Configuration of UrbanSim: In Detail



Paul Waddell, 2011

UrbanSim: Add Relocation Dynamics

Household Location Models

Household Transition Model

Household Relocation Model

Household Location Choice Model

Employment Location Models

Employment Transition Model

Employment Relocation Model

Employment Location Choice Model

Being used in Research Triangle Park, North Carolina. No representation of supply side of real estate market, or prices. Last resort when there is no data on supply.

Paul Waddell, 2011

UrbanSim: Add Real Estate Supply and Price

Land Development Models

- **Real Estate Price Model**
- **Residential Development Project Location Choice Model**
- **Nonresidential Development Project Location Choice Model**
- **Building Construction Model**

Household Location Models

- **Household Transition Model**
- **Household Relocation Model**
- **Household Location Choice Model**

Employment Location Models

- **Employment Transition Model**
- **Employment Relocation Model**
- **Employment Location Choice Model**

Being used in Maricopa County (MAG). Two different levels of geography being tested: travel model zones, and 'super-parcels' or aggregated parcels by land use/block.

Paul Waddell, 2011

UrbanSim: Add Labor Market & Workplace

Land Development Models

- **Real Estate Price Model**
- **Residential Development Project Location Choice Model**
- **Nonresidential Development Project Location Choice Model**
- **Building Construction Model**

Household Location Models

- **Household Transition Model**
- **Household Relocation Model**
- **Household Location Choice Model**

Employment Location Models

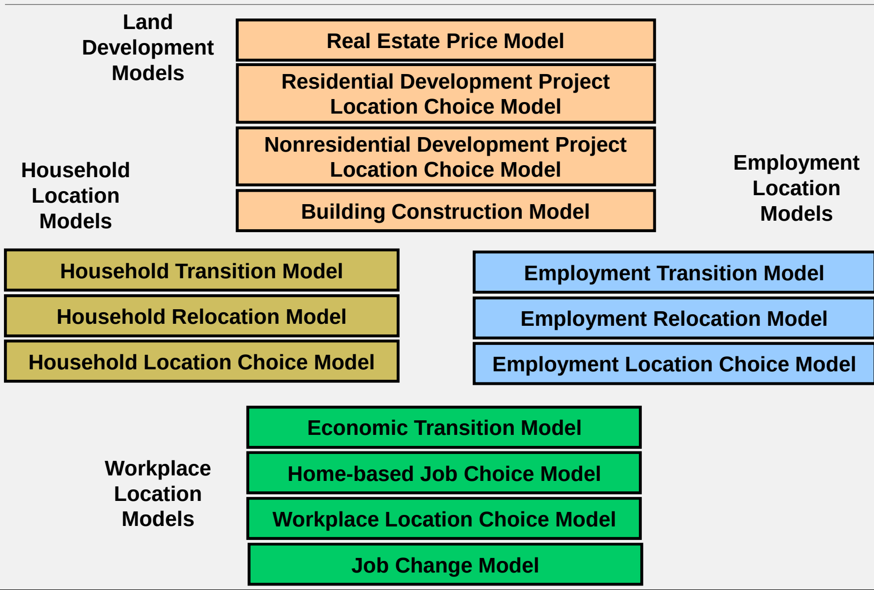
- **Employment Transition Model**
- **Employment Relocation Model**
- **Employment Location Choice Model**

Workplace Location Models

- **Economic Transition Model**
- **Home-based Job Choice Model**
- **Workplace Location Choice Model**
- **Job Change Model**

Paul Waddell, 2011

UrbanSim: Shift From Zones to Parcels as Locations



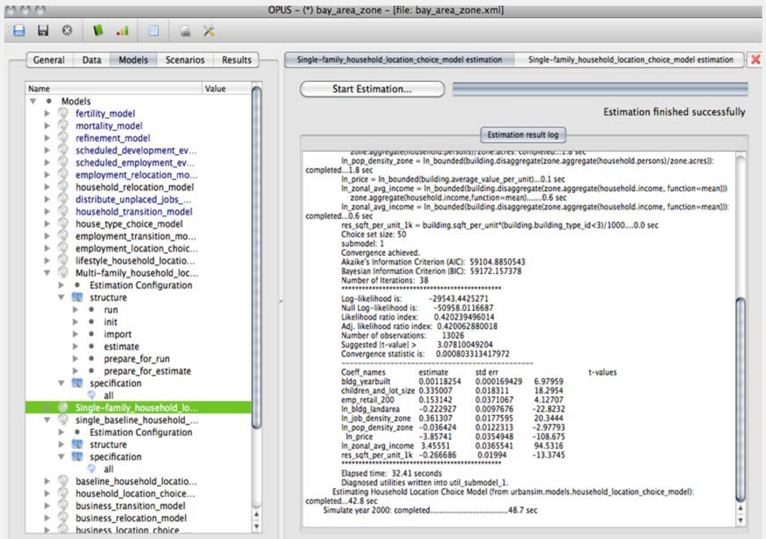
Paul Waddell, 2011

Measuring Progress: Benefits and Costs

- Incremental model development ideally will monitor an evaluate progress:
  - How much better is a new innovation compared to the best available version?
  - How much more costly is it in terms of computational expense or data effort?
- How to measure progress
  - Model estimation results
  - Model sensitivities
  - Calibration of model uncertainty over time

Paul Waddell, 2011

Estimation of Location Choice Models in UrbanSim



Paul Waddell, 2011

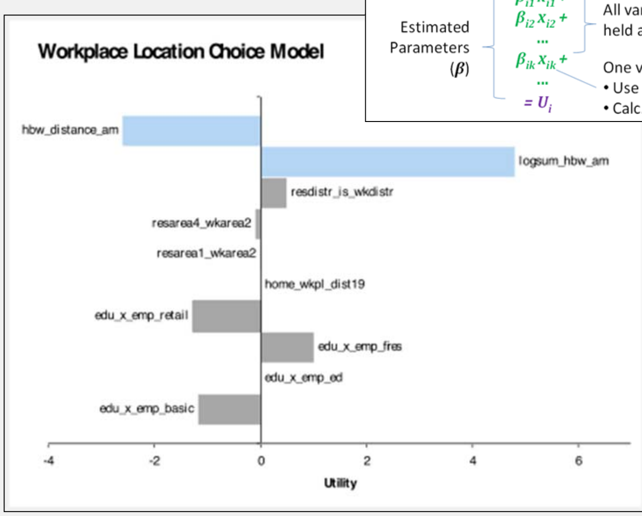
Sensitivity Analysis: Relative Influence of Variables

Estimated Parameters (β)

$\beta_{i1}x_{i1}+$
$\beta_{i2}x_{i2}+$
...
$\beta_{ik}x_{ik}+$
...
$=U_i$

All variables (x) except one held at median value

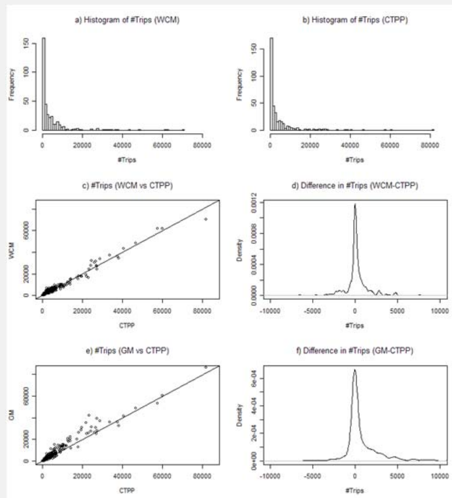
One variable:
- Use 5th & 95th percentile values
- Calc. Δ U for indication of influence



P. Waddell, 2010

Calibration and Validation of Workplace Choice Model

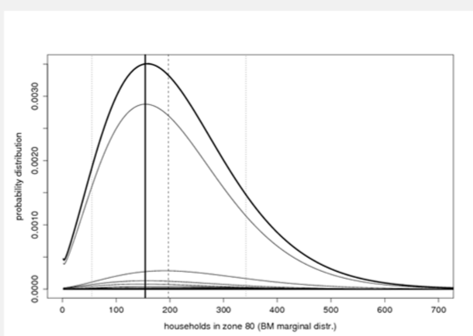
- Model predicts individual workplace (attaches a job to a person) at parcel and building level
- Comparison of predicted values (commute trips) to observed values from CTPP to compute errors
- Calibration: Adding variables to specification to reduce errors.
- Errors compared to previous HBW Trip Distribution Model (gravity model)
  - RMSE Gravity Model = 2558.65
  - RMSE New Model = 1440.01



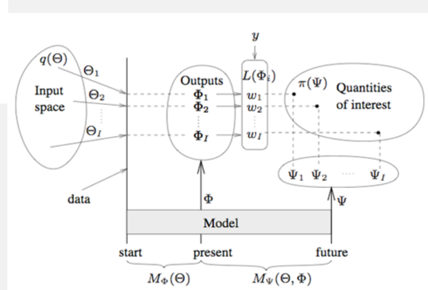
Application: Puget Sound Regional Council

Paul Waddell, 2011

Calibrating Uncertainty in UrbanSim Model Application



Method	# Cases Missed by 90% Confidence Interval	Percent Covered Cases
Bayesian Melding	31	88%
Multiple Runs	163	38%



Bayesian Melding

Results from Eugene-springfield in Transportation Research B, 2007; Seattle Viaduct application in Transportation Research A, 2011.

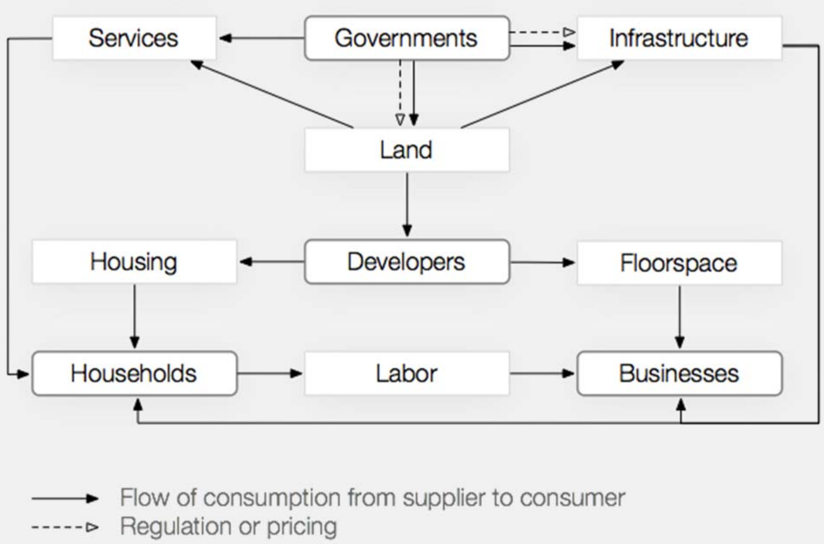
Paul Waddell, 2011

Webinar 6: Modeling Real Estate Demand

1. How Much Detail Do We Need in Models: An Incremental Approach
2. **Disaggregate Location Choice Modeling: A Discrete Choice Approach**
3. Case Studies in Model Development

a) Housing Tenure and Location, Zone Level, San Francisco Bay Area Household Location Choices at Parcel/Building Level, Puget Sound Region
b) Modeling Employment Location Choices

Role of Urban Real Estate Markets



Real Estate Demand: Theoretical Framework

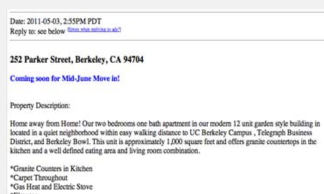
- Macroeconomic Growth
- Product Differentiation
- Market Segmentation
- Market Conditions
- Location Factors
- Disaggregate Demand
- Submarkets
- Demand and Supply Interaction
- Policy Influence

- *Please refer to Webinar 2 for detailed discussion of these topics*

Paul Waddell, 2011

Disaggregate Demand for Real Estate

- All households and firms must choose a location (everyone is an expert in this!)
- Demand is very discrete: consumption is typically a single property
- Real estate (properties) can be described as a bundle of attributes
  - Cannot purchase components 'a la carte' (kitchen, yard, accessibility)
  - Location is part of the bundle
  - Each location is unique, making real estate properties 'quasi-unique'
- The consumer's problem is to choose a single property from those available
- This problem lends itself to representation as a discrete choice model
- Most applicable theory for this is Random Utility Maximization (RUM)



Date: 2011-05-03, 2:55PM PDT
Reply to: see below

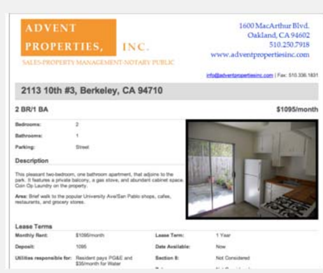
252 Parker Street, Berkeley, CA 94704

Coming soon for Mid-June Move in!

Property Description:

Home away from Home! Our two bedrooms one bath apartment in our modern 12 unit garden style building in located in a quiet neighborhood within easy walking distance to UC Berkeley Campus , Telegraph Business District, and Berkeley Bowl. This unit is approximately 1,000 square feet and offers granite countertops in the kitchen and a well defined eating area and living room combination.

*Granite Counters in Kitchen
*Carpet Throughout
*Gas Heat and Electric Stove
*Elevator



ADVENT PROPERTIES, INC.
SALES-PROPERTY MANAGEMENT-NOTARY PUBLIC

1600 MacArthur Blvd.
Oakland, CA 94602
510.250.7918
www.adventpropertiesinc.com

2113 10th #3, Berkeley, CA 94710

2 BR/1 BA **$1095/month**

Bedrooms: 2
Bathrooms: 1
Parking: Street

Description

This pleasant two-bedroom, one bathroom apartment, that adjoins to the park. It features a private balcony, a gas stove, and abundant cabinet space. Coin Op Laundry on the property.

Area: Brief walk to the popular University Ave/San Pablo shops, cafes, restaurants, and grocery stores.

Lease Terms

Monthly Rent:	$1095/month	Lease Term:	1 Year
Deposit:	1095	Date Available:	Now
Utilities responsible for:	Resident pays PG&E and $35/month for Water	Section 8:	Not Considered

Paul Waddell, 2011

What factors do you consider in looking for housing?

Price? Square footage? Housing type?

Number of bedrooms? Architecture? Floorplan?

School quality? Distance to school/work?

Distance to retail and restaurants? Walkability?

Proximity to transit? Parking? Open space?

Proximity to parks? Neighborhood safety?

Neighborhood demographics?

Paul Waddell, 2011

How do you weigh these factors, make trade-offs?

Price? Square footage? Housing type?

Number of bedrooms? Architecture? Floorplan?

School quality? Distance to school/work?

Distance to retail and restaurants? Walkability?

Proximity to transit? Parking? Open space?

Proximity to parks? Neighborhood safety?

Neighborhood demographics?

Paul Waddell, 2011

Random Utility Maximization

- **Probabilistic** approach; alt. i being chosen:

$$P(i) = P\left[U_i = \max_j U_j\right]$$

- **Utility function**: deterministic & random parts
- $U_i = V_i + \epsilon_i$ where $V_i = \beta \cdot x_i$

- **Multinomial Logit (MNL)** formulation:
  Assume ϵ_i independently & identically distributed with a Gumbel distribution

$$P(i) = \frac{e^{V_i}}{\sum_j e^{V_j}}$$

Multinomial Logit Model of Location Choice

- Most common technique for analyzing disaggregate consumer choice is the multinomial logit model:

$$P_{UL}^{C} = \frac{e^{\alpha(Y-P_U)+\beta S_U+\delta L_U}}{\sum_U e^{\alpha(Y-P_V)+\beta S_V+\delta L_V}}$$

Where
P^C_{UL} is the probability of a consumer of type C choosing site u from the set U
Y is household income or firm profit
P is he price of the building
S is a set of structural characteristics of the building
L is a set of location characteristics at the site are sets of estimated demand parameters

Paul Waddell, 2011

Residential Location Choice

- **Decision makers**: households (HH)

- **Alternatives**: varying degrees of aggregation
  - Zones
  - Neighbourhoods
  - Gridcells (e.g., 150m x 150m)
  - Parcels
  - Buildings
  - Units

Paul Waddell, 2011

Compared to Mode Choice Models...

The choice set is huge!

Paul Waddell, 2011

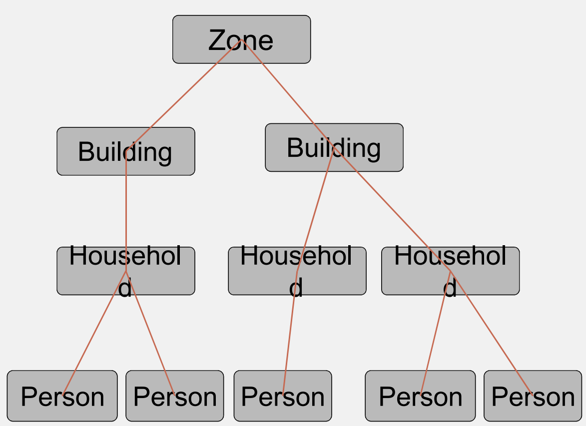
How do we deal with the size of the choice set?

- **In a multinomial logit formulation, random sampling of alternatives will produce consistent, though not efficient estimates of coefficients.**

- However, choice set formation remains a thorny issue.
- Not all randomly sampled alternatives are realistic
- Some alternatives may be “dominated” by others

Paul Waddell, 2011

Zone-Level Data structure in UrbanSim



Attributes

acres
non-res/res acres

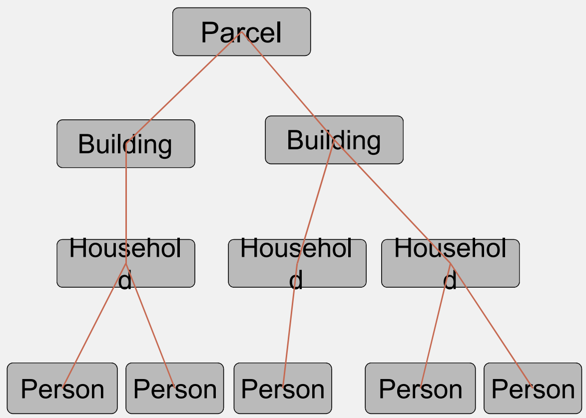
Lot size
Building type
Year built

persons, # cars,
Income, household type

age, workzone,
earnings

Paul Waddell, 2011

Parcel-Level Data structure in UrbanSim



Attributes

acres
non-res/res acres

Lot size
Building type
Year built

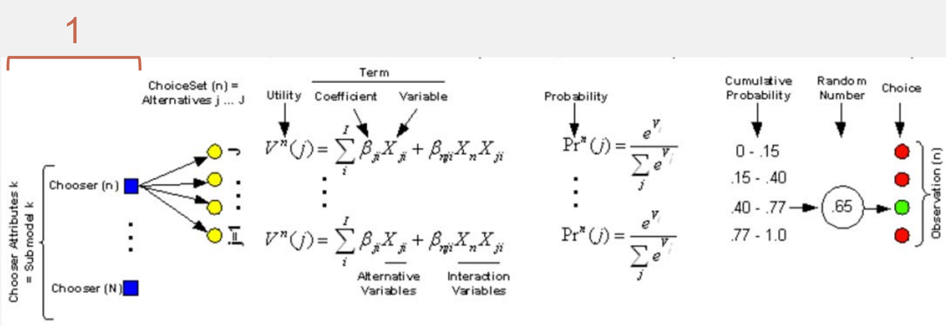
persons, # cars,
Income, household type

age, workzone,
earnings

Paul Waddell, 2011

Urbansim household location choice model

1. Determine households looking for houses (HouseholdTransitionModel and HouseholdRelocationModel)
2. Identify all available housing units within the simulation year.
- Randomly sample alternatives to construct a reasonably-sized choice set.
- Generates location choice probabilities for each household using MNL model
- Used random number sampling to determine which of the available locations each household will choose.
- Once a household has chosen a location, that unit becomes unavailable.

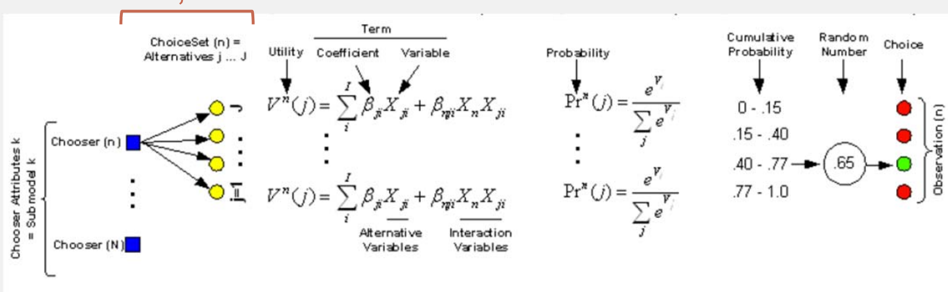


Paul Waddell, 2011

Urbansim household location choice model

1. Determine households looking for houses (HouseholdTransitionModel and HouseholdRelocationModel)
2. Identify all available housing units within the simulation year.
- Randomly sample alternatives to construct a reasonably-sized choice set.
- Generates location choice probabilities for each household using MNL model
- Used random number sampling to determine which of the available locations each household will choose.
- Once a household has chosen a location, that unit becomes unavailable.

2,3

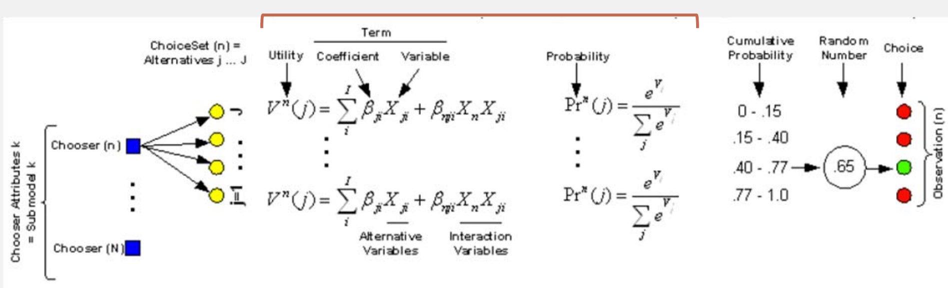


Paul Waddell, 2011

Urbansim household location choice model

1. Determine households looking for houses (HouseholdTransitionModel and HouseholdRelocationModel)
2. Identify all available housing units within the simulation year.
- Randomly sample alternatives to construct a reasonably-sized choice set.
- Generates location choice probabilities for each household using MNL model
- Used random number sampling to determine which of the available locations each household will choose.
- Once a household has chosen a location, that unit becomes unavailable.

4

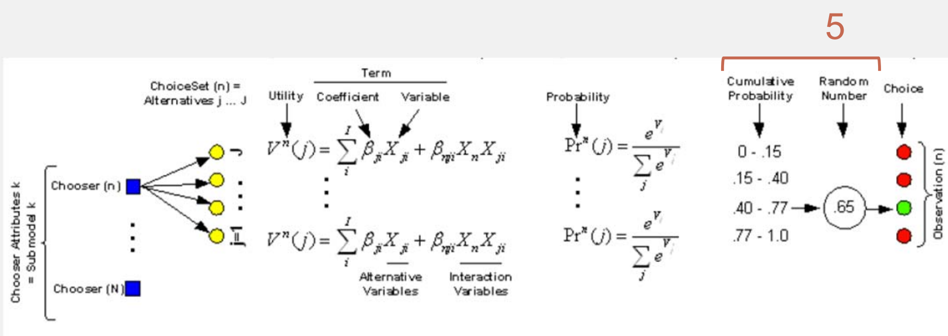


Paul Waddell, 2011

Urbansim household location choice model

1. Determine households looking for houses (HouseholdTransitionModel and HouseholdRelocationModel)
2. Identify all available housing units within the simulation year.

- Randomly sample alternatives to construct a reasonably-sized choice set.
- Generates location choice probabilities for each household using MNL model
- Used random number sampling to determine which of the available locations each household will choose.
- Once a household has chosen a location, that unit becomes unavailable.

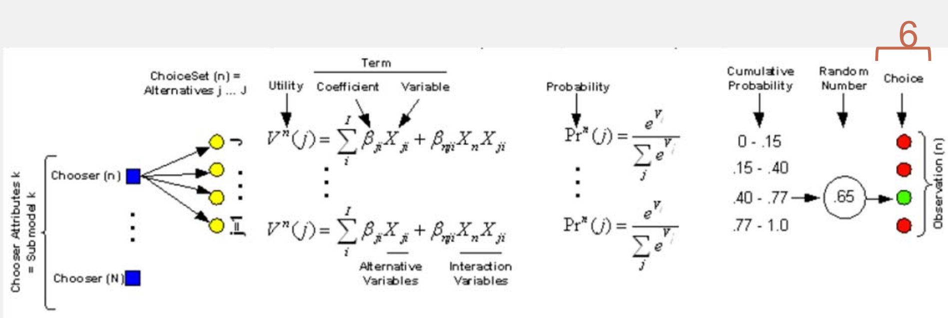


Paul Waddell, 2011

Urbansim household location choice model

1. Determine households looking for houses (HouseholdTransitionModel and HouseholdRelocationModel)
2. Identify all available housing units within the simulation year.

- Randomly sample alternatives to construct a reasonably-sized choice set.
- Generates location choice probabilities for each household using MNL model
- Used random number sampling to determine which of the available locations each household will choose.
- Once a household has chosen a location, that unit becomes unavailable.



Paul Waddell, 2011

Aggregate Demand From Disaggregate Choice Model

$$D = \sum_k P_k(i) = \sum_k \left(\frac{e^{V_i}}{\sum_j e^{V_j}} \right)$$

- where k indexes households and i is a location from the set of J locations

- Compare housing to other choices using the discrete choice framework
  - Choice of transport mode
    - Aggregate demand is the sum of P(i) for each mode (i) * Number of trips
    - Equals total number of trips by mode
  - Choice of housing, at the house level
    - What is aggregate demand in this case?

Paul Waddell, 2011

MNL Models: Pros & Cons

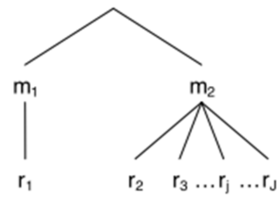
- Closed & tractable mathematical form
- Allows random sampling to estimate model with consistent parameters

- Assumption: Independence from Irrelevant Alternatives (IIA)
– Likely to be violated in housing market
– Nearby houses may share unobserved attributes

Nested Logit Model of Mobility & Relocation Choices

Mobility choice (m)
m_1 = stay
m_2 = move

Relocation choice (r)
r_1 = stay location
r_j = move locations, j = 2, .., J



Observations (N = 4,730)	r_1	$r_2, .., r_{30}$
Stayers ($N_{stayers}$ = 3,062)	chosen alt.	randomly sampled
Movers (N_{movers} = 1,677)	previous location	r_2 = chosen alt. $r_3, .., r_{30}$ = randomly sampled

Lee, Brian and Paul Waddell (2010) Residential Mobility and Location Choice A Nested Logit Model with Sampling of Alternatives, *Transportation*, Volume 37, Issue 4 (2010), pp 587-601.

Other Challenges in Residential Location Choice Models

- Constraints on availability
  - Attractive neighborhoods (central Paris)
  - Prices do not fully ration the market
  - Biases price effect downwards
  - Requires novel estimation methods
- Omitted variable - price endogeneity
  - Omitted variable correlated with quality
  - Will also be correlated with price
  - Biases price coefficient downwards
  - Control function method corrects this

de Palma, A., N. Picard, P. Waddell (2007) Discrete Choice Models with Capacity Constraints: An Empirical Analysis of the Housing Market of the Greater Paris Region. Journal of Urban Economics Vol. 62 (204-230).

Cristian Guevara and Moshe Ben-Akiva (2010) Endogeneity in Residential Location Choice Models, Transportation Research Record: Journal of the Transportation Research Board, No. 1977, Transportation Research Board of the National Academies, Washington, D.C., 2006, pp. 60–66.

Case Study 1: Modeling Choice of Housing Type and Zone-level Location

- Setting: San Francisco Bay Area
- Michelle Thong
- UC Berkeley, Dept. of City and Regional Planning
- Based on graduate courses at UC Berkeley with Paul Waddell and Joan Walker

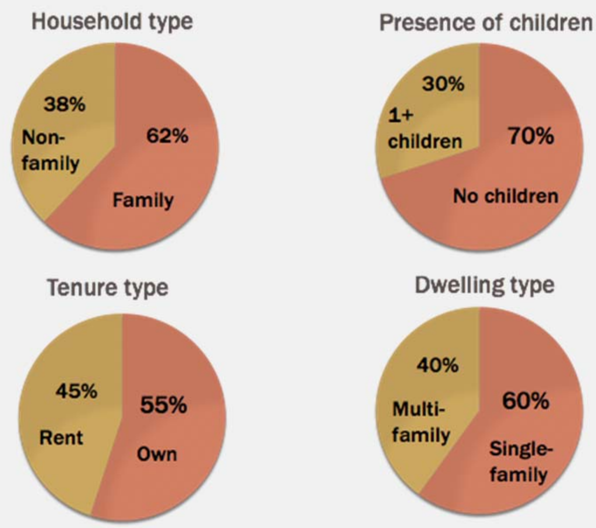
The Household Dataset

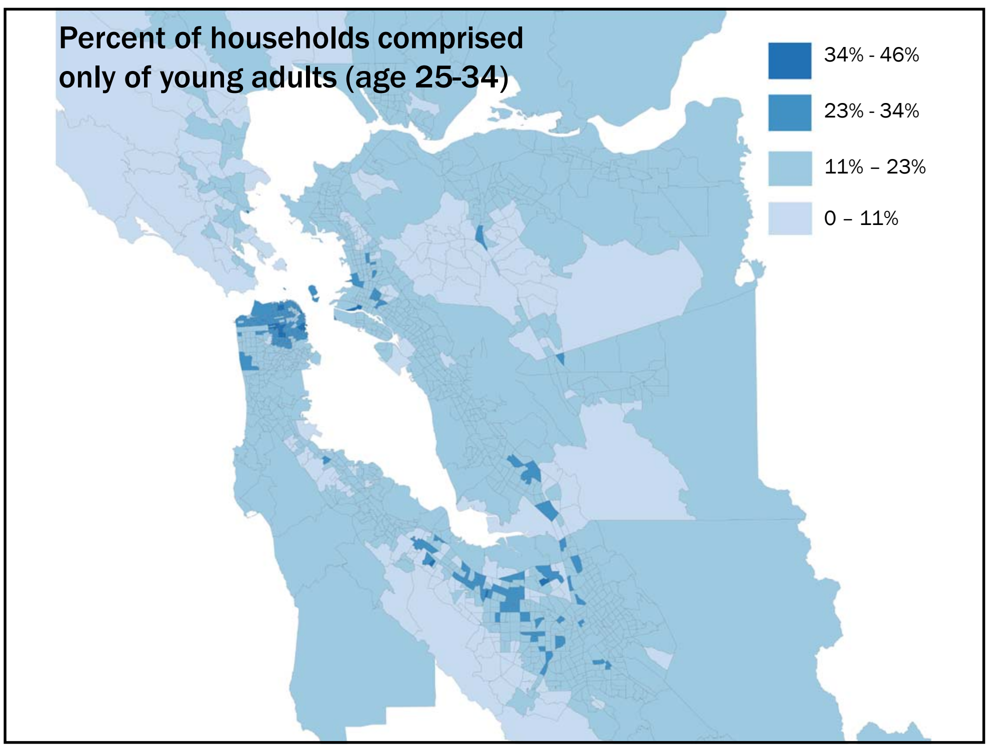
- Consists of a *synthesized* population of households
- Based on data from the 2000 Census and Public Use Microdata sample
- Synthesizer creates a household table and a persons table that is representative of the original dataset
- Each entry has a geographic identifier based on census block group
- Post-processing tools enable synthetic households to be assigned to specific buildings

Paul Waddell, 2011

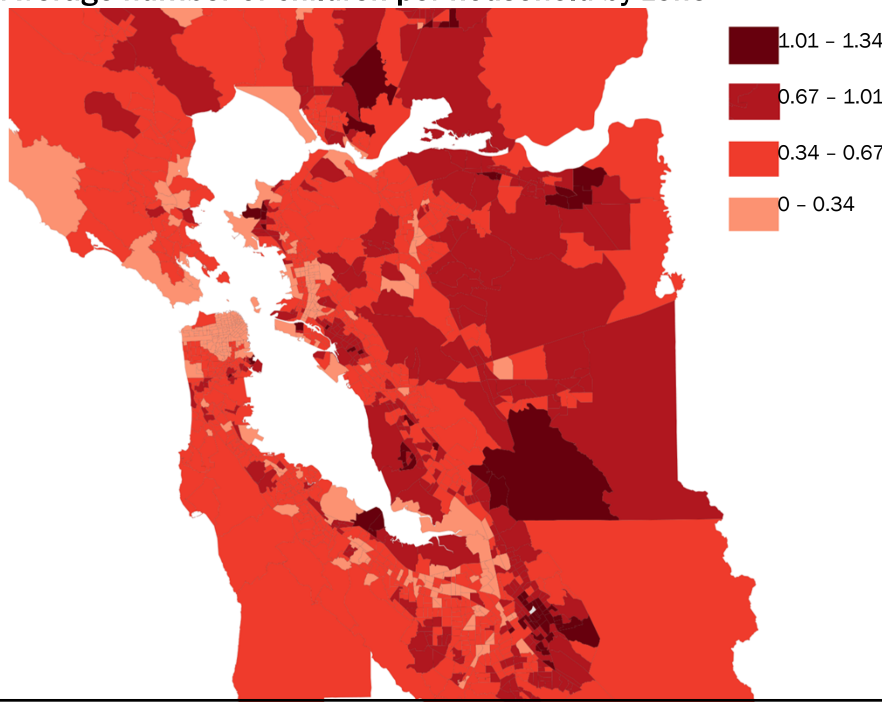
Summary of Household Dataset

- 2.6 million households





Average number of children per household by zone



Paul Waddell, 2011

Limitations of the Dataset

- Household characteristics
- Race/ethnicity

- Attributes of alternatives
- Proximity to and quality of public amenities
- Quality of nearby schools
- Perceived neighborhood safety
- Tenure type
- Neighborhood classification

- Zone-level accessibility measures are not as rich as parcel-level accessibility measures

Paul Waddell, 2011

BASELINE MODEL

Adjusted likelihood ratio index = 0.30

Name	Value	t-statistic
Is_single_family	0.89	34.5
ln(zone job density)	0.19	11.7
ln(zone pop density)	0.085	8.5
ln(price)	-1.08	-135.4
ln(zone avg income)	0.31	13.3
ln(square footage)	1.21	128.2

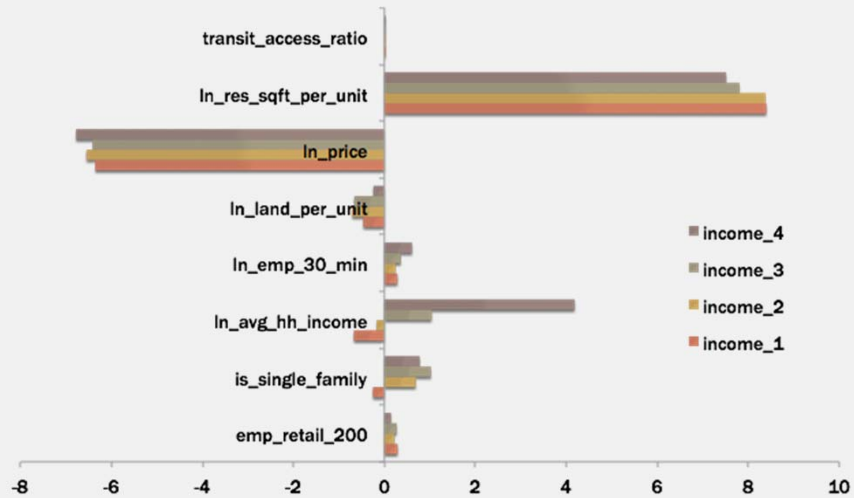
Paul Waddell, 2011

Baseline model stratified by income level

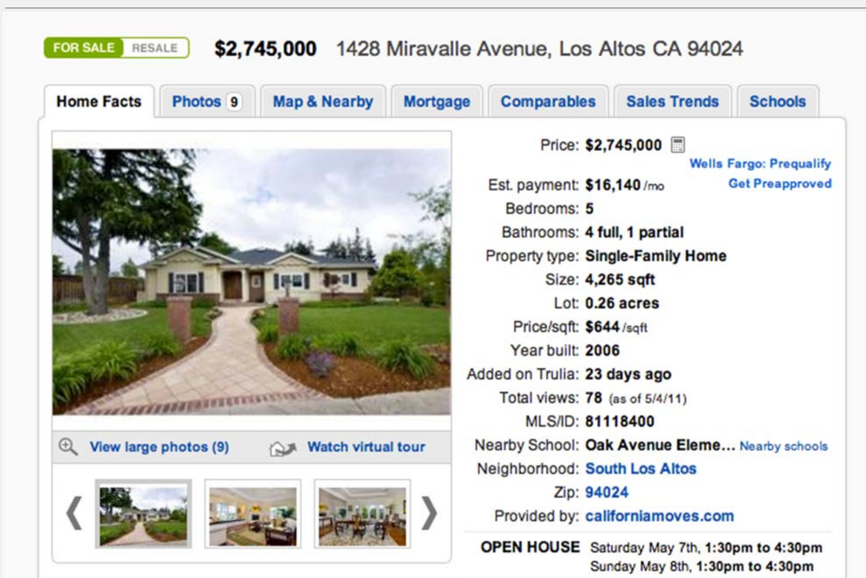
Submodel name	Subgroup description	Adj likelihood ratio index
income_1	Household income between 0 to $20,000	0.27
income_2	Household income between $20,000 to $50,000	0.26
income_3	Household income between $50,000 to $100,000	0.33
income_4	Household income of $100,000 and up	0.45

Baseline model stratified by income level

Differences in utilities between 5th and 95th percentile values of variables

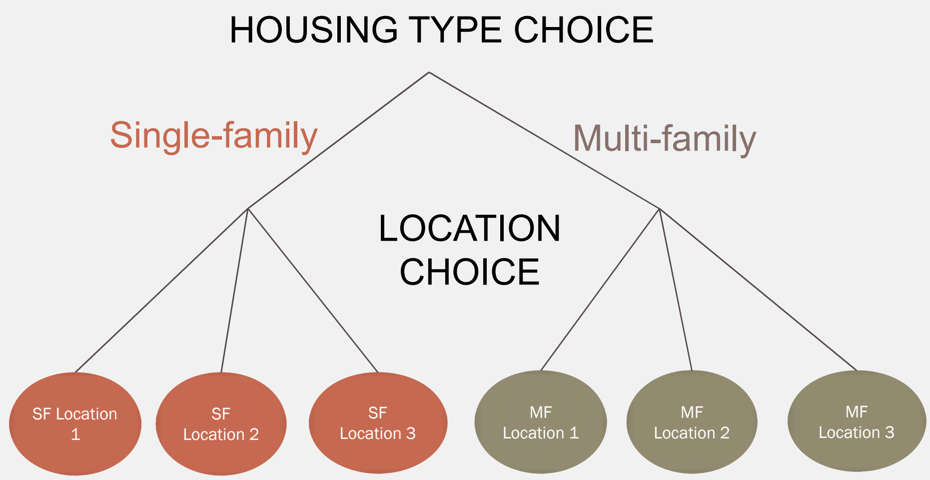


Back to the question of choice set



Paul Waddell, 2011

Nested (Sequential) Model



Paul Waddell, 2011

Housing Type Choice Model (Multi-Family, Single-Family)

Adjusted likelihood ratio index = 0.39

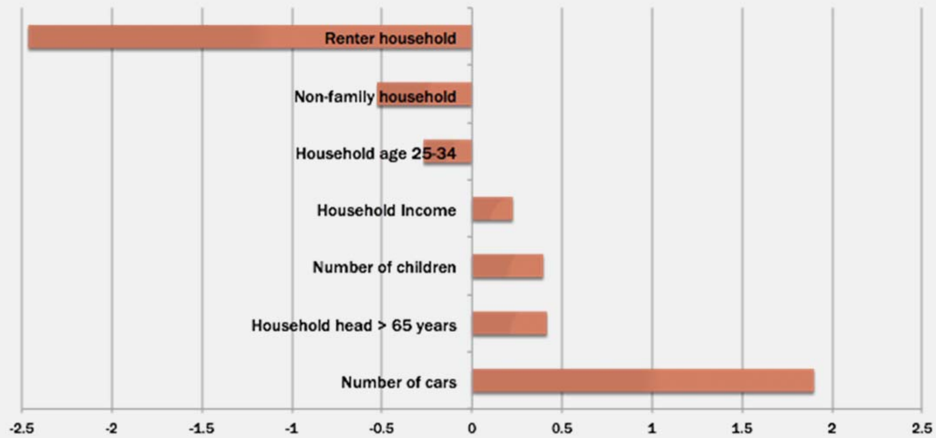
Name	Value	t-statistic
Renter household	-2.46	-672.8
Non-family household	-0.53	-128.2
Household aged 25-34	-0.27	-78.3
Household income	1.41e-6	57.0
# of children	0.20	98.3
Household head > 65 yrs	0.42	84.6
# of cars	0.63	341.9

Omitted alternative is multi-family, so coefficients relate to choice of single-family housing type

Paul Waddell, 2011

Housing type choice model

Differences in utilities between 5th and 95th percentile values of variables



Paul Waddell, 2011

Multi-family household location choice model

Adjusted likelihood ratio index = 0.46

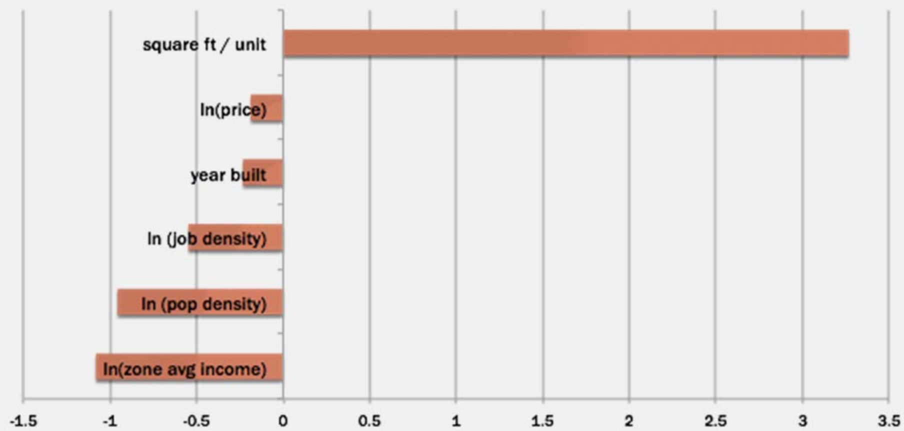
Name	Value	t-statistic
Square footage	3.27	209.3
ln (price)	-0.034	-6.99
Year built	-0.003	-134
ln (zone job density)	-0.123	-10.4
ln (pop density)	-0.231	-20.2
ln (zone avg income)	-0.758	-27.4

Choice is location (zone), using sampling of alternatives. Choice of zone is conditioned on the household having chosen multi-family housing type

Paul Waddell, 2011

Multi-family household location choice model

Differences in utilities between 5th and 95th percentile values of variables



Paul Waddell, 2011

Single-family Household Location Choice Model

Adjusted likelihood ratio index = 0.42

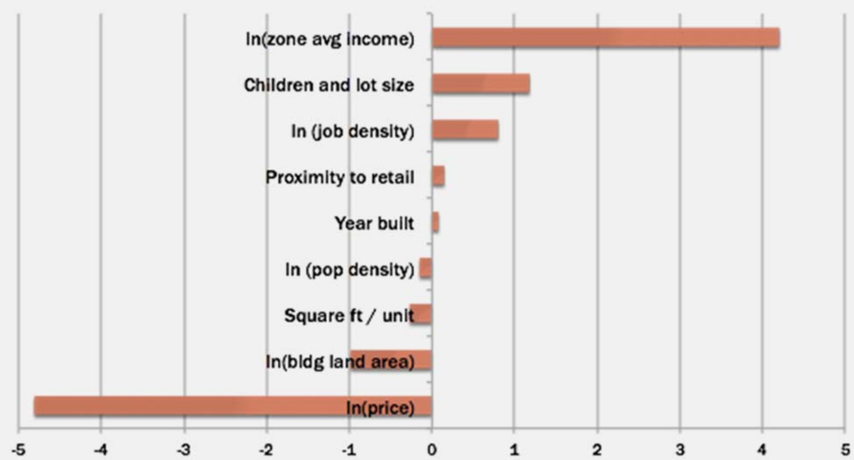
Name	Value	t-statistic
ln (zone avg income)	3.46	94.5
Children x lot size	-0.34	18.3
ln (job density)	0.36	20.3
Proximity to retail	0.15	4.13
Year built	0.001	6.98
ln (pop density)	-0.036	-2.98
Square footage	-0.27	-13.4
ln (bldg land area)	-0.22	-22.8
ln (price)	-3.86	-108.7

Choice is location (zone), using sampling of alternatives. Choice of zone is conditioned on the household having chosen single-family housing type

Paul Waddell, 2011

Single-family household location choice model

Differences in utilities between 5th and 95th percentile values of variables



Paul Waddell, 2011

Housing Demand and Demographic Shifts

- Housing supply likely to shift based on changing demographics
- and expected demand

Household type	1970	2010	2030
With children	45%	33%	27%
Without children	55%	67%	73%

Source: Arthur C. Nelson, Presidential Professor & Director of Metropolitan Research, University of Utah

Modeling Considerations

- There is a discrepancy between demand and supply for different housing types.
- If models are estimated based on existing choices, how do we capture latent demand for more dense forms of housing?

Source: Arthur C. Nelson, Presidential Professor & Director of Metropolitan Research, University of Utah

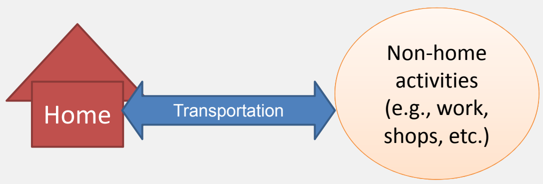
Case Study 2: Parcel-level Household Location

- Setting: Puget Sound Region
- Lee, Brian, Paul Waddell, Liming Wang and Ram Pendyala (2010) Re-examining the Influence of Work and Non-work Accessibility on Residential Location Choices with a Micro-analytic Framework. Environment and Planning A, Volume 42, issue 4, pp 913-930.

Paul Waddell, 2011

Residential Location and Accessibility

Accessibility concept:
ties land use & transportation



Measurements vary in complexity & ability to capture concept

Paul Waddell, 2011

How Much Does Accessibility Matter?

- Access to jobs for households with 0 or 2+ workers?
- Non-work accessibility?
- Trip-chaining? (Tours?)

- **Accessibility no longer important?**
- (Giuliano & Small 1993; Gordon & Richardson 1995)
  - Multiple-worker HH & dispersion of employment
  - Competition with access to other amenities

Application: Parcel/Building Level Location Choice

Central Puget Sound Application

4-county metropolitan Seattle region

16,300 km^2 (6,290 mi^2)

3.5+ mil. people

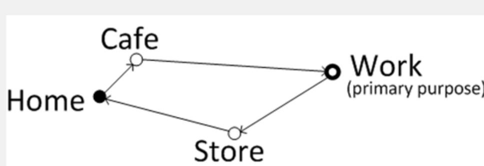
Datasets

Socio-economic, land, travel, geo-spatial

2006 Household Activity Survey

2006 HH Activity Survey

- 2-day activity/travel survey
- 4,739 HH & 10,516 persons
  - Model: subset of recent movers for temporal consistency (1,677 HH)
- Current & past home & workplace locations
- Tour/trip-chain (e.g., home-based work tour)



Paul Waddell, 2011

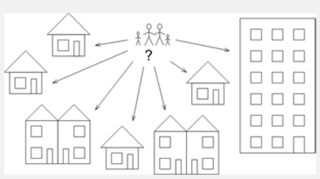
2006 HH Activity Survey Tours

- Descriptive statistics:
- Person-tours: 27,306
- Work tours: 9,257 (34%)
  - 50.1% at least 1 stop other than work
  - 20.2% made stop before work
  - 41.9% made stop after work
  - Majority of stops for shopping, eating out, & personal business

Paul Waddell, 2011

Puget Sound Region
Residential Location Choice Model

- Discrete choice, random utility maximization
- Multinomial Logit formulation



- HH level decisions
- Building level alternatives

- Alternative attributes (e.g., building specific qualities for control; accessibilities)

Paul Waddell, 2011

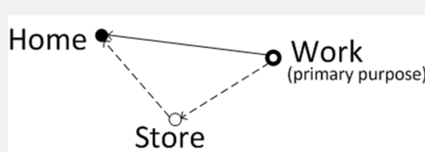
Accessibility - Generalized vs Individual

- **General accessibility**
  – Only place-specific (home-based)
  – Travel costs (e.g., travel time to CBD)
  – Opportunities (e.g., shops within X distance)
- **Individual-specific accessibility**
  – Place- & person-specific (e.g., home-workplace travel time)

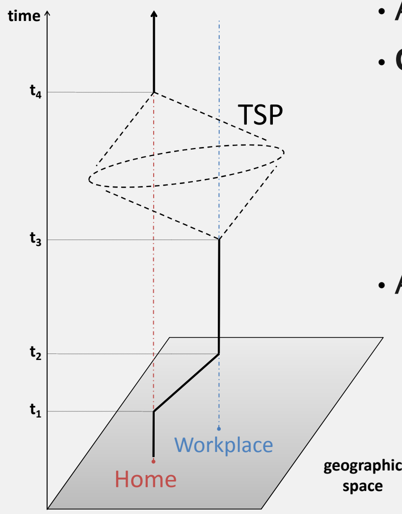
Paul Waddell, 2011

Time-Space Prism Approach

- **Time geography** (Hägerstrand 1970; Miller 1999)
- Explicitly recognizes **time-space constraints**
- Compliments **activity-based** models
- Considers trip-chaining: access to **discretionary** activities between **mandatory** activities



Time-Space Prism (TSP)



- At individual worker level
- **Constraints**
  - Spatial: home & workplace
  - Temporal: home & work schedules
  - Travel: transportation network & travel speeds
- Application: **work-to-home trips**
  - Used survey departure & arrival times for those who made >1 stop (joint distribution)
  - Randomly assigned to each worker

TSP-based Accessibility

- Accessible zones
  - Identify set of zones that can be reached
  - Based on spatial & temporal constraints
  - Use travel speeds from network travel model
- Accessibility opportunity set
  - Number of consumer type jobs (in retail, food & other services sectors)

Explanatory variables

Variables	Descriptions
residential_units	Log of number of residential units in building
same_area_type	Dummy of building in same area type as previous household (HH) location
same_area	Dummy of building in same area as previous HH location
Kitsap	Dummy of building in Kitsap County
population_density	Log of zonal population density
high_inc_x_size	High HH income (inc) dummy x log of average dwelling size (sq ft/unit)
mid_inc_x_size	Mid HH inc dummy x log of average dwelling size (sq ft/unit)
low_inc_x_size	Low HH inc dummy x log of average dwelling size (sq ft/unit)
disposable_inc	log of HH inc less average dwelling price per unit
inc_x_condo	Log of HH inc x building is condo residential dummy
inc_x_MFR	Log of HH inc x building is multi-family residential dummy
kids_x_SFR	Dummy of HH with children x is single-family residential dummy
kids_x_kids_HH	Dummy of HH with children x percent HH with children within walking distance (600m)
one_pers_x_not_SFR	One person HH dummy x building is not single-family residential dummy
renter_x_is_MFR	Renter HH dummy x building is multi-family residential dummy
young_x_young_HH	Young HH (average adult age ≤ 30) dummy x percent young HH within walking distance (600m)
Accessibility variables	
average_logsum	Trip weighted zonal average logsum for AM home-based-work (HBW) drive alone trips
neigh_shopping	Log of number of shopping type jobs (retail, food, & other services) within walking distance
work_travel_time	Travel time for AM HBW drive alone trips to workplace (maximum between up to 2 workers)
TSP_shopping	Log of number of shopping type jobs (retail, food, & other services) in time-space prism for the work-to-home leg of the HBW tour (maximum between up to 2 workers)

Estimation results

	(1) Base Model		Accessibility Models							
			(2) General		(3) Neighborhood		(4) Workplace		(5) Work-to-home	
Variables	coeff.	t-val.	coeff.	t-val.	coeff.	t-val.	coeff.	t-val.	coeff.	t-val.
residential_units	0.674	18.0	0.721	17.6	0.692	17.4	0.629	16.1	0.638	16.2
same_area_type	0.414	4.11	0.417	4.12	0.303	2.87	0.349	3.40	0.380	3.80
same_area	2.65	26.3	2.66	26.4	2.80	27.0	2.34	23.3	2.53	24.3
Kitsap	0.916	9.68	0.830	8.39	0.866	8.50	1.76	14.7	1.73	14.4
population_density	0.0165	0.661	-0.0495	-1.49	-0.101	-2.90	-0.199	-5.48	-0.247	-6.77
high_inc_x_size	0.971	11.6	0.952	11.3	1.01	11.9	0.816	9.67	0.892	10.5
mid_inc_x_size	-0.330	-8.85	-0.333	-8.24	-0.304	-7.55	-0.313	-7.86	-0.311	-7.30
low_inc_x_size	-0.306	-5.37	-0.335	-6.46	-0.320	-5.50	-0.291	-4.67	-0.330	-5.49
disposable_inc	0.0631	3.20	0.0492	2.44	0.0657	3.23	0.0672	3.16	0.0714	3.31
inc_x_condo	0.0539	3.11	0.0405	2.17	0.0479	2.52	0.0524	2.81	0.0591	3.25
inc_x_MFR	-0.117	-6.27	-0.126	-6.54	-0.127	-6.52	-0.136	-6.90	-0.134	-6.91
kids_x_SFR	0.565	2.80	0.647	3.21	0.743	3.53	0.518	2.46	0.610	2.93
kids_x_kids_HH	0.0139	2.85	0.0120	2.45	0.0186	3.73	0.0262	5.13	0.0212	4.20
one_pers_x_not_SFR	0.688	4.33	0.575	3.40	0.581	3.46	0.773	4.70	0.740	4.37
renter_x_is_MFR	2.85	14.6	2.90	14.5	3.05	15.0	2.84	14.0	3.00	14.7
young_x_young_HH	0.0211	5.18	0.0247	5.55	0.0206	4.82	0.0220	4.98	0.0235	5.38
Accessibility variables										
average_logsum			0.635	3.06	0.573	2.88	0.711	3.61	0.808	3.84
neigh_shopping					0.0715	3.88	0.0566	3.05	0.0507	2.70
work_travel_time							-0.0248	-21.1	-0.0227	-17.6
TSP_shopping									0.0473	3.37
Log-likelihood (LL)		-3954.9		-3934.0		-3898.2		-3697.5		-3642.0
Adj. likelihood ratio (ρ')		0.30381		0.30730		0.31341		0.34841		0.35797
χ^2 likelihood test		n/a		Reject model (1) at >99.9% confid.		Reject model (2) at >99.9% confid.		Reject model (3) at >99.9% confid.		Reject model (4) at >99.9% confid.

N = 1677; Null log-likelihood = -5703.8

Estimation Result Highlights

Variables	(1) Base	Accessibility Models (2) Regional work coeff.	t-val.	(3) Neigh. shops coeff.	t-val.	(4) Indiv. work TT coeff.	t-val.	(5) Indiv. TSP shop ops. coeff.	t-val.
Control variables									
…									
Accessibility variables									
average_logsum		0.64	3.1	0.57	2.9	0.71	3.6	0.81	3.8
neigh_shopping				0.072	3.9	0.057	3.1	0.051	2.7
work_travel_time						-0.025	-21	-0.023	-18
TSP_shopping								0.048	3.4
Log-likelihood (LL)	-3954.9	-3934.0		-3898.2		-3697.5		-3642.0	
Adj. likelihood ratio (ρ’)	0.30381	0.30730		0.31341		0.34841		0.35797	
χ^2 likelihood test	n/a	Reject (1) at >99.9%		Reject (2) at >99.9%		Reject (3) at >99.9%		Reject (4) at >99.9%	

N = 1677; Null log-likelihood = -5703.8

Relative Influence of Variables

- Use estimated parameters & utility function

Estimated Parameters ($\boldsymbol{\beta}$)

alt. i

$$\beta_{i1}x_{i1} + \beta_{i2}x_{i2} + \dots \beta_{ik}x_{ik} + \dots = U_i$$

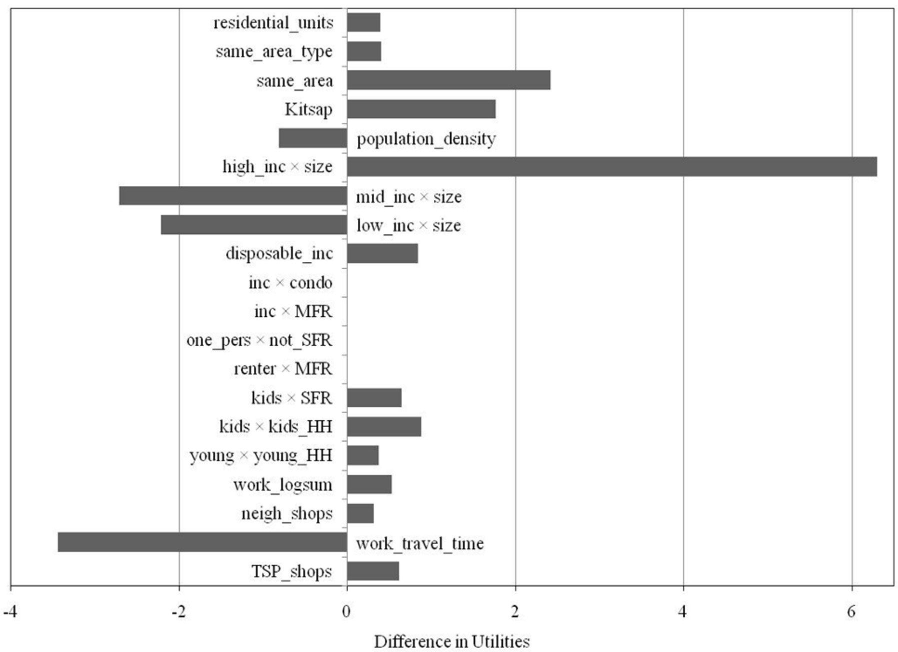
All variables ($\boldsymbol{x}$) except one held at median value

One variable:
- Use 5th & 95th percentile values
- Calc. Δ $\boldsymbol{U}$ for indication of influence

- (HH income) x (building size): greatest influence
- Individual work travel time: 2nd overall
- Individual TSP shop ops.: 2nd highest accessibility

Relative Influence on Utility of Residential Location:

Difference in utilities between the 5th and 95th percentile values



Conclusions

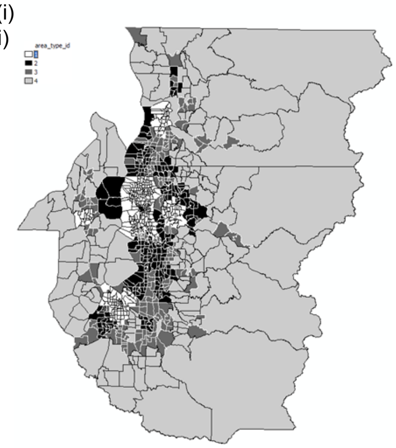
- **Accessibility** still matters for residential choice
  - Work & discretionary activities
- **Time-space prism** constrained approach
  - Theoretically sound for accessibility measurement
  - Captures trip-chaining discretionary activities
- **Disaggregate** approach
  - Captures omitted variables
  - Activity-based micro-simulation

Prediction-Success Table

By Transportation Analysis Zones (TAZ) area types

Row(i) = observed number of households in category (i)
Col(i) = predicted number of households in category (i)
Cell(i,i) = matrix diagonal
Success rate(i) = Cell(i,i) / Sum of Row(i)
Predicted/Observed = Sum of Col(i) / Sum of Row(i)

Figure 2: TAZ area types



	Success	Predicted			
Observed	Rate	1	2	3	4
1	0.528	383	137	132	74
2	0.262	127	97	79	68
3	0.286	113	88	102	54
4	0.166	65	54	67	37
Predicted/Observed		0.948	1.014	1.064	1.045

(1 = metropolitan cities, 2 = 2 core suburban cities,
3 = smaller suburban cities, 4 = rural areas)

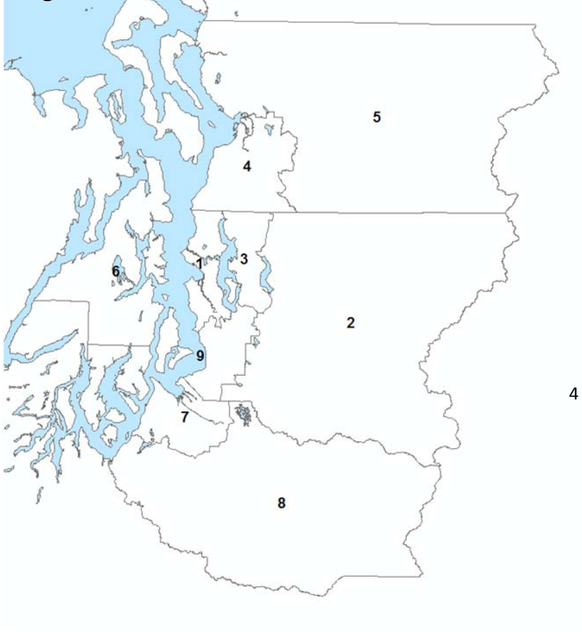
Prediction-Success Table

By Forecast Analysis Zone (FAZ) districts (aggregation of TAZ)

	Success	Predicted								
Observed	Rate	1	2	3	4	5	6	7	8	9
1	0.527	300	32	41	57	11	37	32	18	41
2	0.239	28	26	12	6	2	12	5	9	9
3	0.173	24	15	19	16	4	6	16	1	9
4	0.259	65	12	28	60	13	21	13	3	17
5	0.095	10	7	4	21	6	7	3	0	5
6	0.327	43	11	11	17	9	66	21	13	11
7	0.226	30	12	9	15	5	26	40	25	15
8	0.178	12	4	6	12	3	5	16	16	16
9	0.128	31	13	9	10	5	16	17	8	16
Predicted/Observed		0.954	1.211	1.264	0.922	0.921	0.970	0.921	1.033	1.112

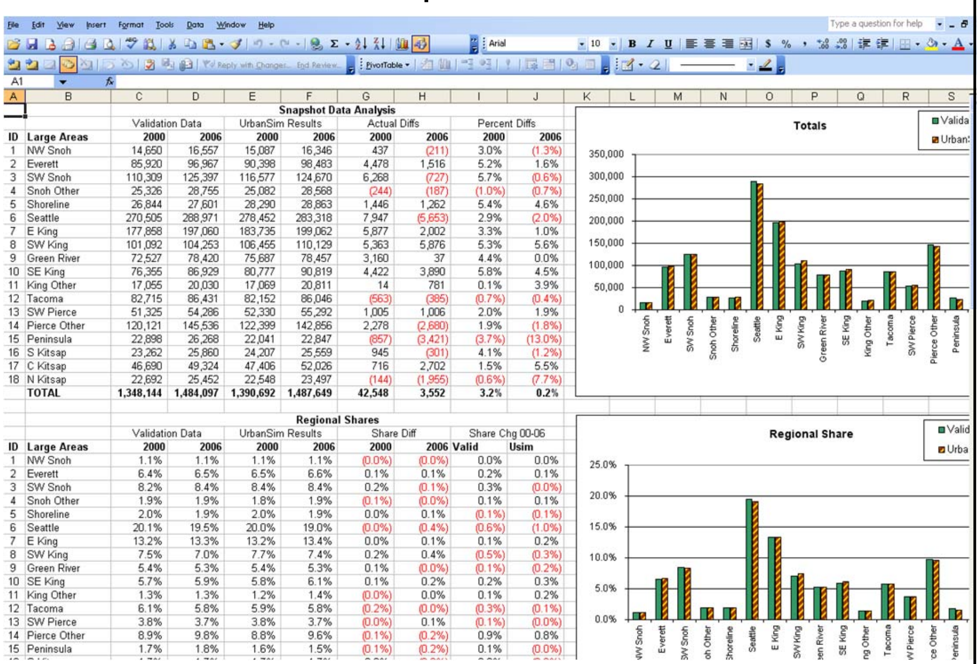
(1 = Seattle, 2 = East King County, 3 = East suburbs in King County,
4 = North suburbs in Snohomish County, 5 = North & east Snohomish County,
6 = Kitsap County, 7 = Tacoma, 8 = Pierce County except Tacoma,
9 = South suburbs in King County)

Figure 3: FAZ districts



1 = Seattle
2 = East King County
3 = East suburbs in King County
4 = North suburbs in Snohomish County
5 = North & east Snohomish County
6 = Kitsap County
7 = Tacoma
8 = Pierce County except Tacoma
9 = South suburbs in King County

Validation over Multiple Years of Simulation



Snapshot Data Analysis

		Validation Data		UrbanSim Results		Actual Diffs		Percent Diffs	
ID	**Large Areas**	**2000**	**2006**	**2000**	**2006**	**2000**	**2006**	**2000**	**2006**
1	NW Snoh	14,650	16,557	15,087	16,346	437	(211)	3.0%	(1.3%)
2	Everett	85,920	96,967	90,398	98,483	4,478	1,516	5.2%	1.6%
3	SW Snoh	110,309	125,397	116,577	124,670	6,268	(727)	5.7%	(0.6%)
4	Snoh Other	25,326	28,755	25,082	28,568	(244)	(187)	(1.0%)	(0.7%)
5	Shoreline	26,844	27,601	28,290	28,863	1,446	1,262	5.4%	4.6%
6	Seattle	270,505	288,971	278,452	283,318	7,947	(5,653)	2.9%	(2.0%)
7	E King	177,858	197,060	183,735	199,062	5,877	2,002	3.3%	1.0%
8	SW King	101,092	104,253	106,455	110,129	5,363	5,876	5.3%	5.6%
9	Green River	72,527	78,420	75,687	78,457	3,160	37	4.4%	0.0%
10	SE King	76,355	86,929	80,777	90,819	4,422	3,890	5.8%	4.5%
11	King Other	17,055	20,030	17,069	20,811	14	781	0.1%	3.9%
12	Tacoma	82,715	86,431	82,152	86,046	(563)	(385)	(0.7%)	(0.4%)
13	SW Pierce	51,325	54,286	52,330	55,292	1,005	1,006	2.0%	1.9%
14	Pierce Other	120,121	145,536	122,399	142,856	2,278	(2,680)	1.9%	(1.8%)
15	Peninsula	22,898	26,268	22,041	22,847	(857)	(3,421)	(3.7%)	(13.0%)
16	S Kitsap	23,262	25,860	24,207	25,559	945	(301)	4.1%	(1.2%)
17	C Kitsap	46,690	49,324	47,406	52,026	716	2,702	1.5%	5.5%
18	N Kitsap	22,692	25,452	22,548	23,497	(144)	(1,955)	(0.6%)	(7.7%)
	TOTAL	**1,348,144**	**1,484,097**	**1,390,692**	**1,487,649**	**42,548**	**3,552**	**3.2%**	**0.2%**

Regional Shares

		Validation Data		UrbanSim Results		Share Diff		Share Chg 00-06	
ID	**Large Areas**	**2000**	**2006**	**2000**	**2006**	**2000**	**2006**	**Valid**	**Usim**
1	NW Snoh	1.1%	1.1%	1.1%	1.1%	(0.0%)	(0.0%)	0.0%	0.0%
2	Everett	6.4%	6.5%	6.5%	6.6%	0.1%	0.1%	0.2%	0.1%
3	SW Snoh	8.2%	8.4%	8.4%	8.4%	0.2%	(0.1%)	0.3%	(0.0%)
4	Snoh Other	1.9%	1.9%	1.8%	1.9%	(0.1%)	(0.0%)	0.1%	0.1%
5	Shoreline	2.0%	1.9%	2.0%	1.9%	0.0%	0.1%	(0.1%)	(0.1%)
6	Seattle	20.1%	19.5%	20.0%	19.0%	(0.0%)	(0.4%)	(0.6%)	(1.0%)
7	E King	13.2%	13.3%	13.2%	13.4%	0.0%	0.1%	0.1%	0.2%
8	SW King	7.5%	7.0%	7.7%	7.4%	0.2%	0.4%	(0.5%)	(0.3%)
9	Green River	5.4%	5.3%	5.4%	5.3%	0.1%	(0.0%)	(0.1%)	(0.2%)
10	SE King	5.7%	5.9%	5.8%	6.1%	0.1%	0.2%	0.2%	0.3%
11	King Other	1.3%	1.3%	1.2%	1.4%	(0.0%)	0.0%	0.1%	0.2%
12	Tacoma	6.1%	5.8%	5.9%	5.8%	(0.2%)	(0.0%)	(0.3%)	(0.1%)
13	SW Pierce	3.8%	3.7%	3.8%	3.7%	(0.0%)	0.1%	(0.1%)	(0.0%)
14	Pierce Other	8.9%	9.8%	8.8%	9.6%	(0.1%)	(0.2%)	0.9%	0.8%
15	Peninsula	1.7%	1.8%	1.6%	1.5%	(0.1%)	(0.2%)	0.1%	(0.0%)

Questions and Discussion

Presenters:
Paul Waddell
Department of City and Regional Planning
University of California Berkeley
email: waddell at berkeley.edu

Mike Reilly
Association of Bay Area Governments