Simple Synthetic Populations Without the Use of Random Draws

Vince Bernardin, Jr., Ph.D.
Bernardin, Lochmueller & Associates, Inc.

Bernardin · Lochmueller & Associates, Inc.

Planners · Engineers · Surveyors

Types of Models

- A new type of population synthesizer for a new type of travel model

	Analytic	Simulation
Aggregate		
Disaggregate		

Types of Models

- A new type of population synthesizer for a new type of travel model

	Analytic	Simulation
Aggregate	Four-step	
Disaggregate		

Types of Models

- A new type of population synthesizer for a new type of travel model

	Analytic	Simulation
Aggregate	Four-step	
Disaggregate		Activity-Based

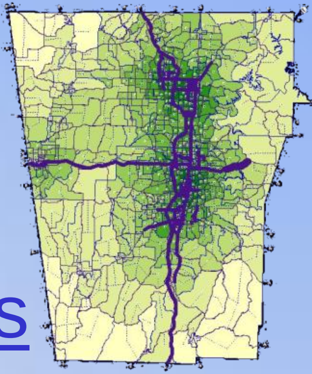
Types of Models

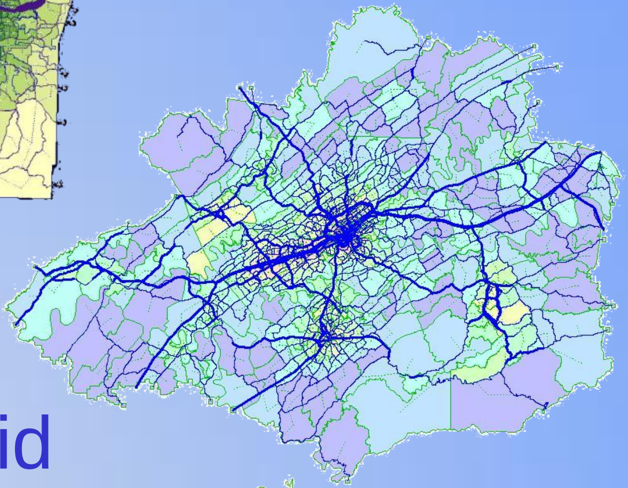
- A new type of population synthesizer for a new type of travel model

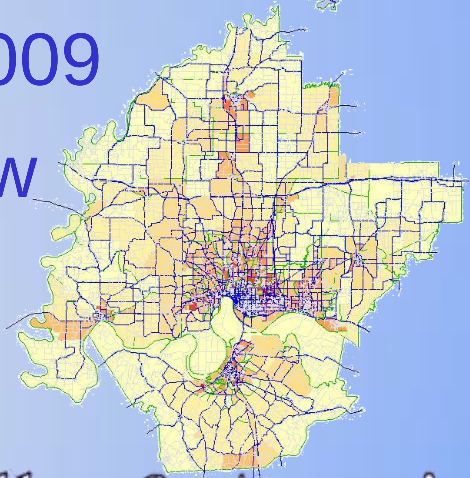
	Analytic	Simulation
Aggregate	Four-step	
Disaggregate	“Hybrid”	Activity-Based

Background

- History
  - Northwest Arkansas Hybrid Aggregate/ Disaggregate, 2006
  - Knoxville Region 1st Hybrid Trip-based/Tour-based, 2009
  - Evansville Metro Area New Hybrid Model, 2011







Why build hybrids?

- Aggregation Bias
  - Consider the probability of transit use for:
    - 100 households with an average of 2.2 cars per household
    - 5 households with no cars, 15 hh with one car, 50 hh with two cars, 20 hh with three cars, 5 hh with four cars, 5 hh with five
  - Non-linear choice probabilities

Why build hybrids?

- Aggregation Bias
- Efficiency
  - Consider the following scenario:
    - A sample of 10-12 activity-based simulations per scenario can be required for corridor-level auto loadings, transit, etc.
    - A single run typically takes 12-24 hours on a machine with 10-15 processors
    - Comparing just 4 alternatives can take a month of computing time
    - We would all do well to take this seriously

The Challenge

- Eliminate aggregation bias
  - Produce a synthetic population
- Without using random draws

Approaches

- Rule-based heuristics
  - Shift people, jobs, income, students, etc., between households until criteria are met
  - Northwest Arkansas
- Weighting
  - Define all possible household categories and weight them by their probability x population
  - Knoxville, Evansville

Weighting

- Advantage: It's easy!
- Limitation: Size
  - Records in population database depend on number of household types & grow quickly with the number of household characteristics
  - But, not as bad as you might think, since many combinations of characteristics are not possible or observed

Knoxville's Synthetic Population

- Households only (not individuals)
- 6 variables
  - Number of Persons: 1-5+
  - Number of Workers: 0-3+
  - Number of Students: 0-2+
  - Presence of Seniors: 0,1
  - Income Group: low, mid, high
  - Vehicle Ownership: 0-4+

Database Size

- Possible Size
  - 5 x 4 x 3 x 2 x 3 x 5 x TAZ = 1800 x TAZ
- Constraints
  - e.g., workers <= persons
  - Observed 157 x 5 x TAZ = 785 x TAZ
- Knoxville
  - ~1,000 TAZ => 800,000 records
  - More than actual number of households
  - But easy to produce and manageable size

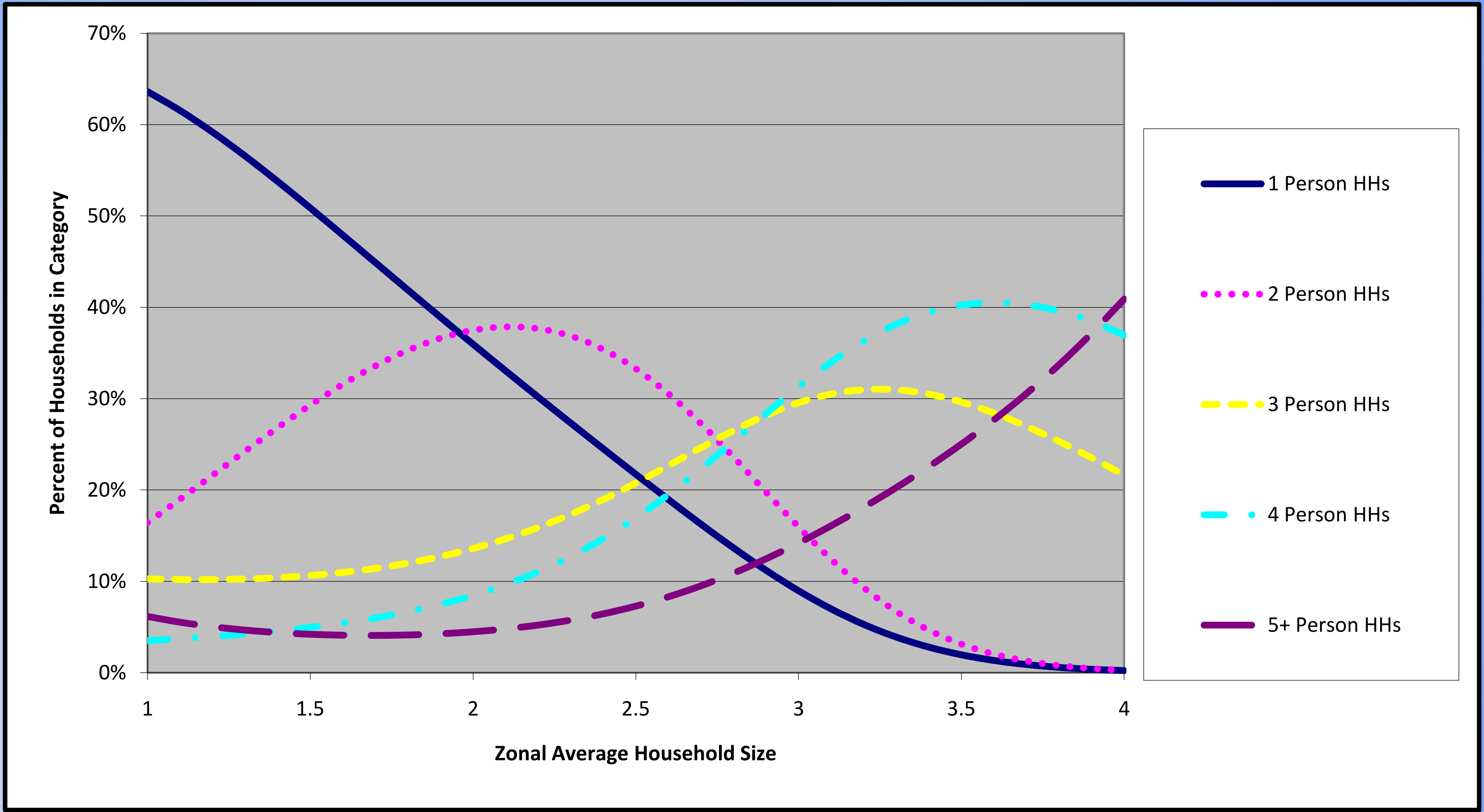
Three Easy Steps

- Univariate, marginal distributions
  - Aggregate ordinal logit models
  - With shadow prices to enforce means
- Combined, multivariate distribution
  - Iterative proportional fitting
- Vehicle Availability / Auto Ownership
  - Disaggregate ordinal logit model

Aggregate Ordinal Logit Models

- OLM outperformed (un-nested) MNL
- Simple, aggregate logit models driven by distribution’s mean

Univariate Marginal Distributions



Aggregate Ordinal Logit Models

- OLM outperformed (un-nested) MNL
- Simple, aggregate logit models driven by distribution’s mean
  - Secondary variables contribute, too
    - e.g., for a given zonal students/hh, zero student households are more likely with seniors

Household Size OLM

Household Size	Alternative	Parameter	t-statistic
-- Logsum Parameters			
Nest_1	alt_2, Nest_2	0.9	Constrained
Nest_2	alt_3, Nest_3	0.8	Constrained
Nest_3	alt_4, alt_5	0.7	Constrained
-- Alternative Specific Parameters			
CONSTANT	alt_1	1.4991	1.15
CONSTANT	alt_2	-4.2750	-2.18
CONSTANT	alt_3	-0.4124	-0.29
CONSTANT	alt_4	-1.9605	-1.35
Zonal Average Household Size	alt_1	2.5378	2.05
Zonal Average Household Size	alt_2	4.9789	2.96
Zonal Average Household Size	alt_3	1.5143	1.26
Zonal Average Household Size	alt_4	1.9344	1.58
Zonal Average Household Size, Squared	alt_1	-0.9999	-3.55
Zonal Average Household Size, Squared	alt_2	-1.3571	-3.70
Zonal Average Household Size, Squared	alt_3	-0.3655	-1.39
Zonal Average Household Size, Squared	alt_4	-0.3655	Constrained
Population Density	alt_1	0.0581	2.07
Log of Zonal Average HH Income	alt_1	-0.3076	-2.41
Log of Zonal Average HH Income	alt_2	0.3827	3.43
Percent of Households with Senior	alt_3	-1.5443	-2.62
-- Model Statistics	statistic		
Log Likelihood at Zero	-4730.5		
Log Likelihood at Constants	-4363.7		
Log Likelihood at Convergence	-4229.1		
Rho Squared w.r.t. Zero	0.106		
Rho Squared w.r.t Constants	0.031		

Shadow Prices

- Used to guarantee output distribution has the mean given as input
- Developed iteratively
  - For alternatives less than the given mean:

$$s_i = s_{i-1} + (TrueAvg - AltAvg)\ln\left(EstAvg_{i-1}/TrueAvg\right)$$

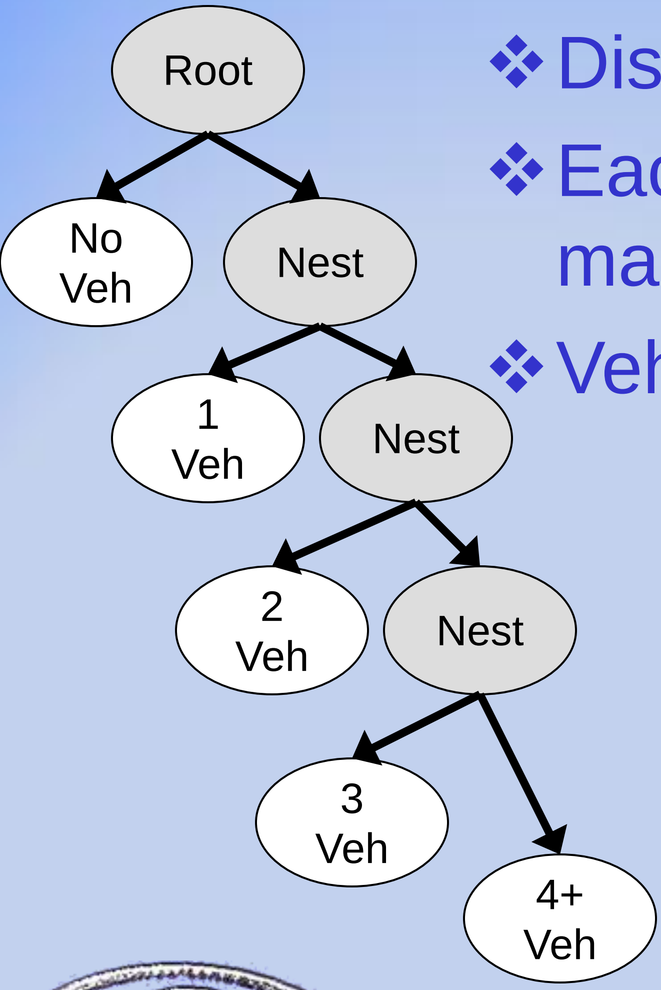
  - For alternatives greater than the given mean:

$$s_i = s_{i-1} + (TrueAvg - AltAvg)\ln\left(TrueAvg/EstAvg_{i-1}\right)$$

Iterative Proportional Fitting

- Marginals from OLM
- Seed distribution of households (from PUMS, etc.)
- Iterative row and column factoring (factors = target/current) converges on distribution with given marginals

Vehicle Availability Choice



- ❖ Disaggregate OLM
- ❖ Each individual household chooses how many vehicles to own / lease
- ❖ Vehicle ownership levels respond to
  - Demographics (household size, income, number of workers, students, etc.)
  - Gas Prices
  - Transit Availability
  - Urban Design (pedestrian environment / grid vs. cul-de-sac design)

Closing Thoughts

- Weighting (using OLM & IPF) is viable for creating disaggregate, analytic "hybrid" models
- Rule-based approach deserves further exploration due to size considerations
- More work needed on the effects of simulation variation on model outputs, especially for transit & disaggregate results

Questions?

VBernardin2@BLAinc.com