Population Synthesis

Activity-Based Microsimulation Model - InSITE

Feb. 2, 2016

Background (cont.)

- Activity-based microsimulation model operates at the level of individual traveler, while recognizing inter dependencies between activities, trips, persons, time and space.

- Activity-based model offers the capability of addressing emerging policy questions of interest by simulating activities and travel of each individual traveler.

Background (cont.)

- The disaggregate data about the socio-economic and demographic characteristics of the entire population in the model region is a fundamental input to the activity-based microsimulation model.

- The detailed disaggregate person and household data for the entire model region is not available.

Objective

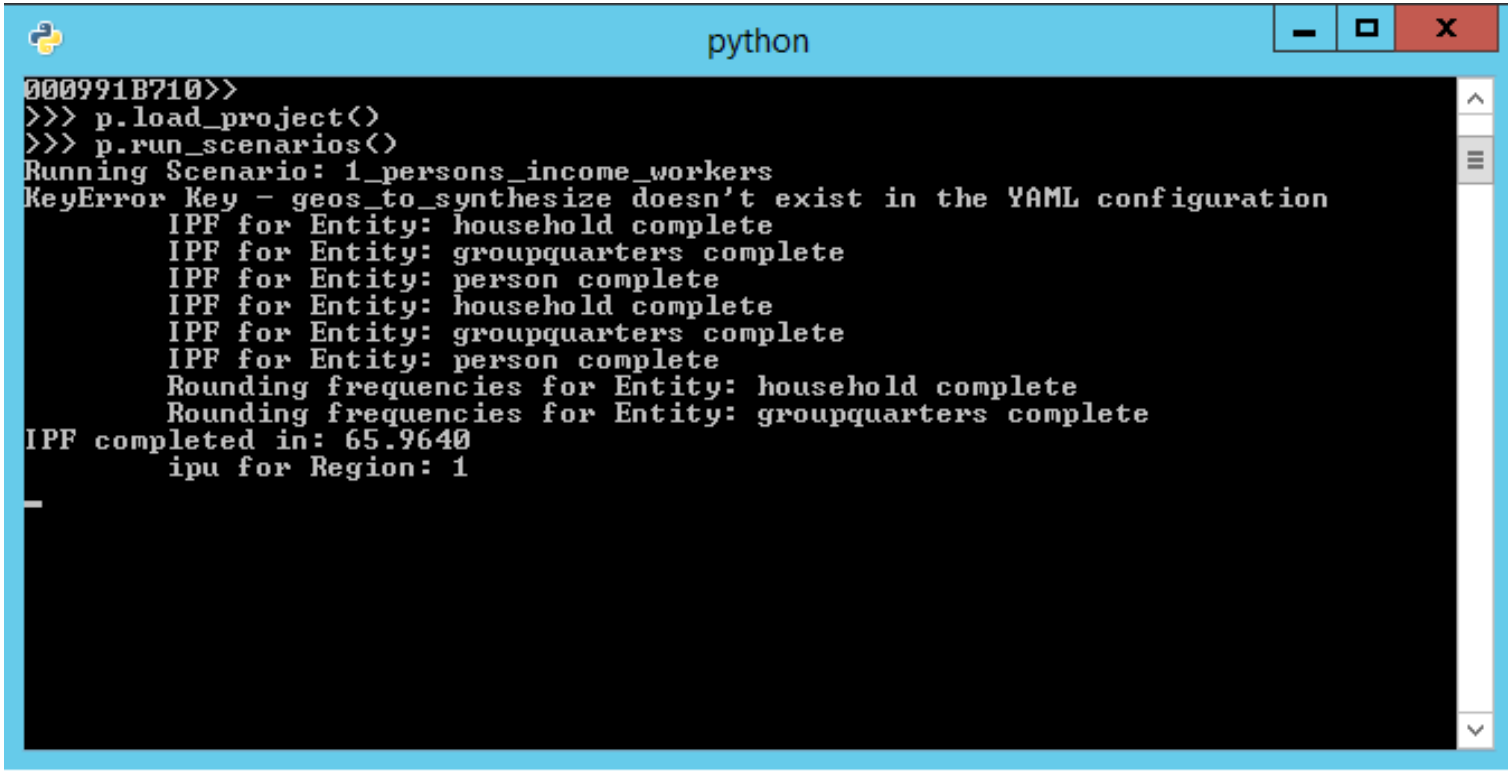
- Generate complete synthetic population with comprehensive information at both person and household levels in the entire BMC region for the activity-based model.

Population Synthesis

- To generate a synthetic population by expanding the household and person sample data to mirror the aggregate marginal distributions of household and person attributes of interest.

PopGen

- Python-based software



PopGen Input

- Household/Person/Groupquarter Sample Data
  - ACS Public Use Microdata Sample (PUMS) data
- Household/Person/Groupquarter Marginal Control Data
  - Aggregate distributions of **control variables** at county level and TAZ level
- Geography Correspondence file
  - Correspondence between county and TAZ

Control Variables

- County Level
  - Household variable: # of workers by household Income
  - Person variable: age by gender
  - Group quarter: number of GQs
- TAZ Level
  - Household variable: HH size, HH Income
  - Person variable: employment status
  - Group quarter: number of IGQ and NIGQ

PopGen Work Steps

- **Step1:** household sample weights are generated that closely match the marginal distribution of household and person control variables.
- **Step2:** estimated sample weights are used to probabilistically draw households (and subsequently the associated person) to generate synthetic population for the given geographic unit.

PopGen Output

- Synthetic household file

- Synthetic person file (including both household persons and group quarter persons)

Validation

- Baltimore City
  - # of workers by household income

Baltimore City		Income Group (2010 $)				
Workers in Household	Total Households	< $15K	$15K - $30K	$30K - $50K	$50K- $100K	$100K+
Total households	0%	0%	0%	0%	0%	0%
0 workers	0%	0%	0%	-1%	1%	-1%
1 worker	0%	0%	0%	0%	0%	0%
2 workers	0%	-1%	-2%	1%	0%	0%
3+ workers	0%	13%	8%	-3%	-1%	1%

Validation

- Baltimore County
  - # of workers by household income

		Income Group (2010 $)				
Workers in Household	Total Households	< $15K	$15K - $30K	$30K - $50K	$50K- $100K	$100K+
Total households	0%	0%	0%	0%	0%	0%
0 workers	0%	0%	-1%	0%	1%	0%
1 worker	0%	-1%	1%	0%	0%	0%
2 workers	0%	-1%	1%	0%	0%	0%
3+ workers	0%	-3%	-3%	-6%	0%	0%

Validation

- Baltimore City
  - Household size by household income

		Income Group (2010 $)				
Persons in Household	Total Households	< $15K	$15K - $30K	$30K - $50K	$50K- $100K	$100K+
Total households	0%	0%	0%	0%	0%	0%
1 person	0%	-4%	-7%	-1%	7%	28%
2 persons	0%	3%	12%	-1%	-2%	-7%
3 persons	0%	0%	-3%	8%	-3%	0%
4 persons	0%	22%	1%	7%	-6%	-8%
5+ perons	0%	10%	7%	-13%	1%	2%

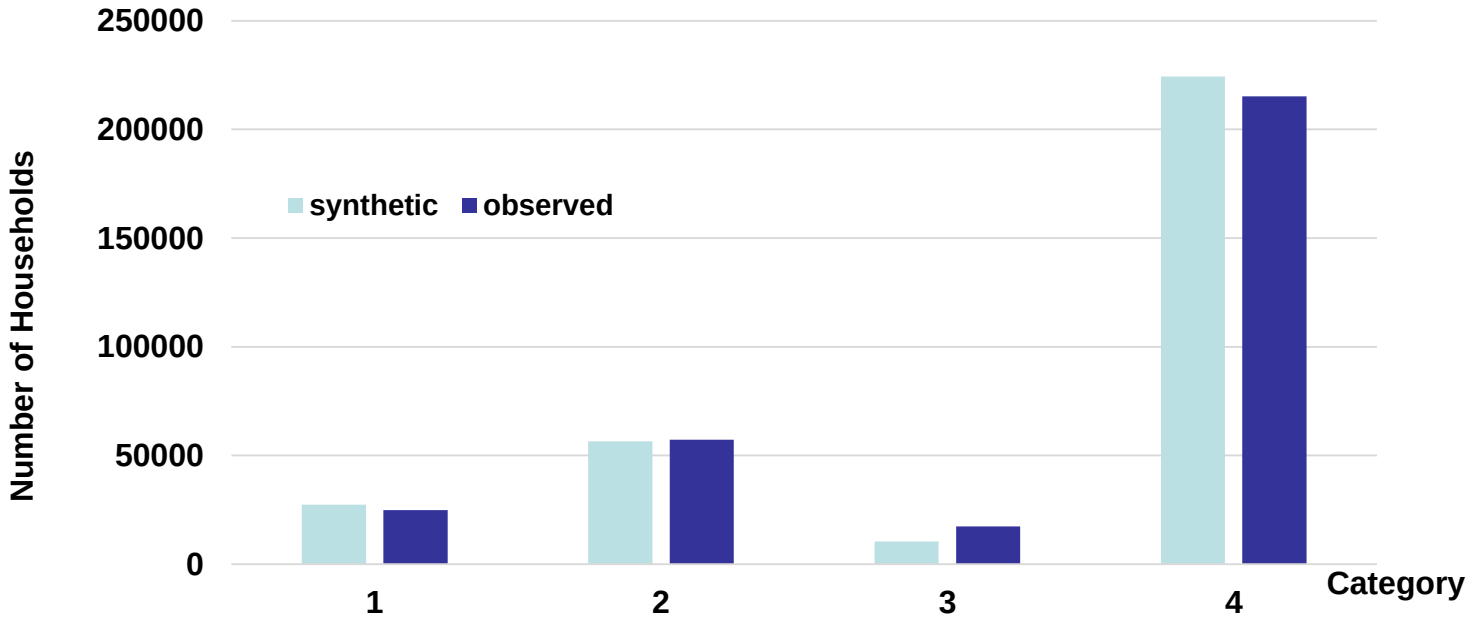
Validation

- Baltimore County
  - Household size by household income

		Income Group (2010 $)				
Persons in Household	Total Households	< $15K	$15K - $30K	$30K - $50K	$50K- $100K	$100K+
Total households	0%	0%	0%	0%	0%	0%
1 person	0%	3%	-2%	2%	-3%	2%
2 persons	0%	-5%	7%	1%	0%	-1%
3 persons	0%	-3%	-5%	-9%	3%	1%
4 persons	0%	4%	-6%	3%	0%	-1%
5+ persons	0%	-25%	-3%	-1%	1%	2%

Validation

- Baltimore County
  - Household with presence of children



1: w/ children under 6 years only; **2:** w/ children 6 to 17 years only
3: w/ children under 6 years and 6 to 17 years; **4:** w/o children

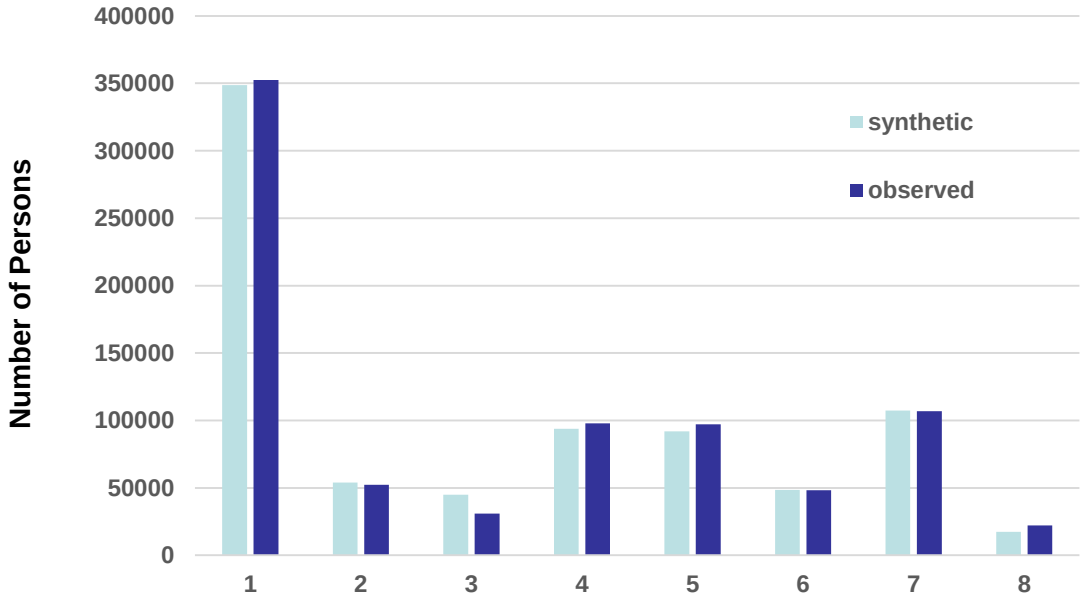
Validation

- Baltimore County
  - Person age (10) by gender (2)

	Age Group				
Gender	**<5**	**5-9**	**10-15**	**16-17**	**18-20**
Male	1%	0%	2%	19%	2%
Female	1%	5%	4%	24%	5%
	21-34	**35-44**	**45-64**	**65-74**	**75+**
Male	3%	0%	1%	0%	2%
Female	5%	0%	1%	0%	0%

Validation

- Baltimore County
  - Persons by person type



Code	Category
1	full time worker
2	part time worker
3	adult student
4	senior
5	non-working adult
6	child 1- age 0-4
7	child 2- age 5-15
8	child 3- age 16-17

Conclusions

- The population synthesizer (PopGen) can generate a TAZ level synthetic population while controlling for marginal distributions at both county level and TAZ level.

- The synthetic population match the observed distributions well for the variables that are important factors in activity-based model.

Conclusions

- It is desirable to control for variables at the lowest level (TAZ level).

- PopGen has difficulty in distinguishing adult student from the non-working adult, as the employment status control variable only controls workers and non-workers.

For More Information

Yijing Lu
410-732-0500 x1052
ylu@baltometro.org
Charles Baber
410-732-0500 x1056
cbaber@baltometro.org
www.baltometro.org