TMIP Webinar Series

Activity-Based Modeling

Session 5: Population Synthesis and Household Evolution

The Travel Model *Improvement* Program

Speakers: John Gliebe & Peter Vovsha

April 26, 2012

Acknowledgments

This presentation was prepared through the collaborative efforts of Resource Systems Group, Inc. and Parsons Brinckerhoff.

- Presenters
  - John Gliebe, Peter Vovsha
- Moderator
  - Stephen Lawe
- Content Development, Review and Editing
  - Bhargava Sana, John Gliebe, Peter Vovsha, John Bowman, Mark Bradley, Joel Freedman, Maren Outwater
- Media Production
  - Brian Grady

2012 Activity-Based Modeling Webinar Series

Executive and Management Sessions	
Executive Perspective	February 2
Institutional Topics for Managers	February 23
Technical Issues for Managers	March 15
Technical Sessions	
Activity-Based Model Frameworks and Techniques	April 5
Population Synthesis and Household Evolution	April 26
Accessibility and Treatment of Space	May 17
Long-Term and Medium Term Mobility Models	June 7
Activity Pattern Generation	June 28
Scheduling and Time of Day Choice	July 19
Tour and Trip Mode, Intermediate Stop Location	August 9
Network Integration	August 30
Forecasting, Performance Measures and Software	September 20

Learning Outcomes

By the end of this session, you will be able to:

- Describe a synthetic population
- Describe the methods used to synthesize a population
- Describe the process of household evolution

Session Outline

- Role of synthetic populations in activity modeling
- Specifying a synthetic population
- Methods used to create a synthetic population
- Synthetic populations in practice
- Household evolution models

Terminology

- Socio-demographic attribute
- Controlled attributes
- Disaggregate household sample
- Seed data
- Iterative Proportional Fitting (IPF)
- Drawing a sample
- Uncontrolled attributes
- Household evolution

Aggregate vs. Disaggregate Travel Representation

Trip-Based Models

Household Cross-Tabulation → Aggregate Zonal Trips

Daily Household Trip Rates → Aggregate Zonal Trips

Activity-Based Models

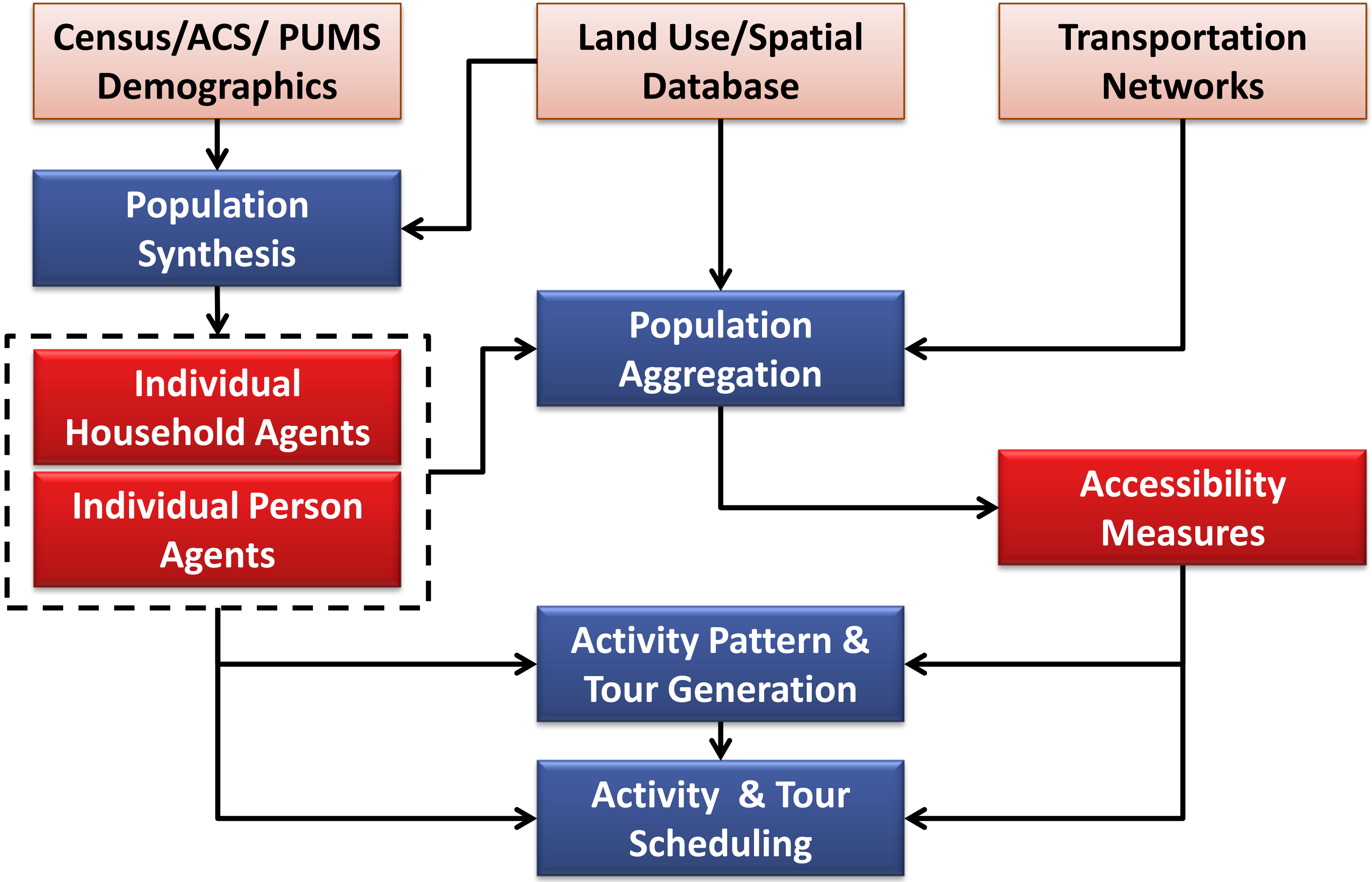
Disaggregate Population → Individual Activity Patterns

Individual Choices → Individual Activity Patterns

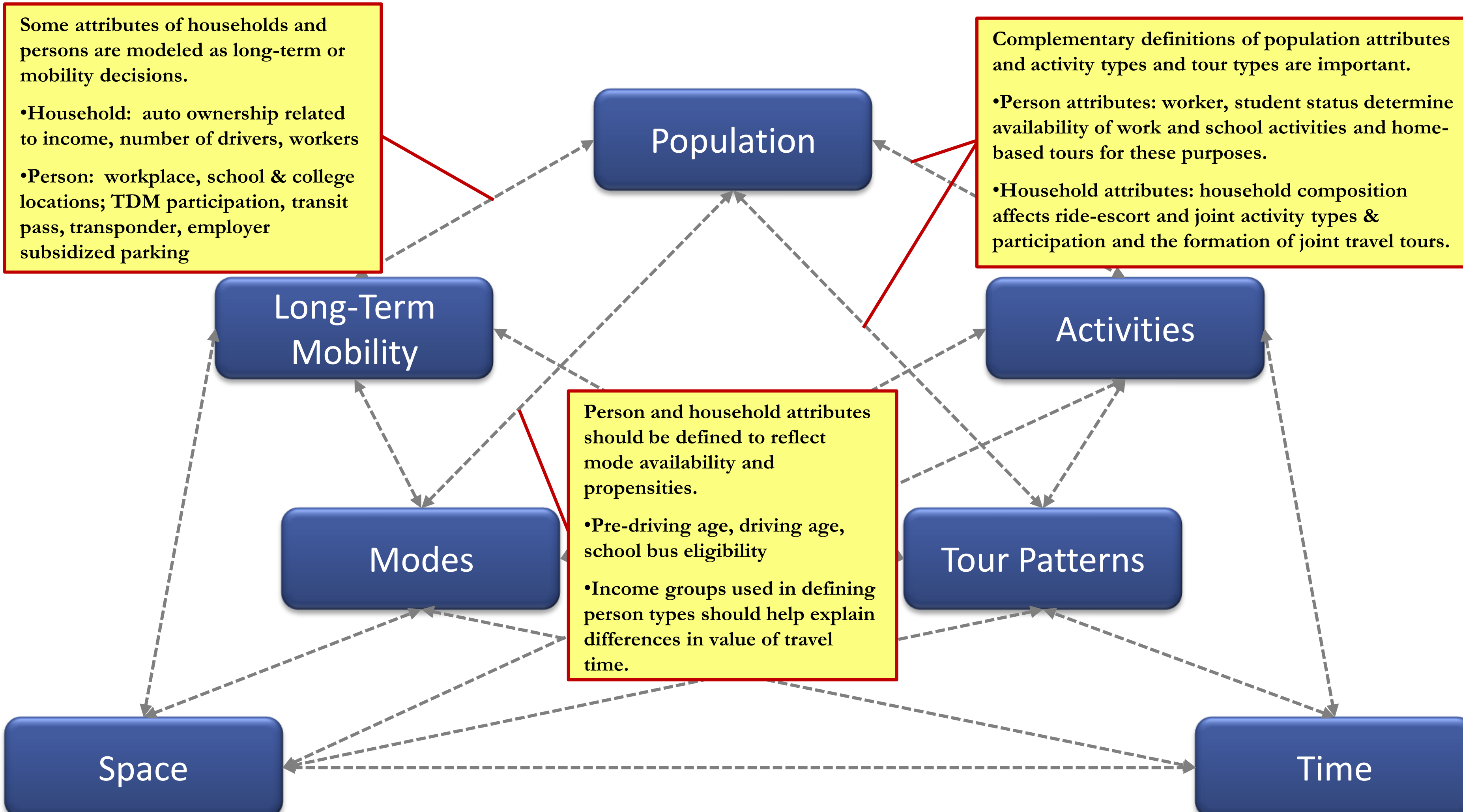
Role of Population Synthesis

- Households and persons are represented individually in the activity-based model through micro-simulation
- Population synthesis "creates" these households and persons for use in the activity-based model
  - Synthetic households and persons should possess all of the demographic attributes needed for model inputs
  - Variables that will be used to explain variation in …
    - Daily activity patterns and tour generation
    - Levels of participation in various activities
    - Preferences for time of day, mode and destination attributes
    - Value of time (willingness to pay)
    - Coordination between household members

Where does population synthesis fit in?



Relations between activity model design elements



Role of Person Types

No.	Person Type	Age	Work Status	School Status
1	Full-time worker	18+	Full-time	None
2	Part-time worker	18+	Part-time	None
3	Non-working adult	18 – 64	Unemployed	None
4	Non-working senior	65+	Unemployed	None
5	College student	18+	Any	College +
6	Driving age student	16-17	Any	Pre-college
7	Non-driving student	6 – 16	None	Pre-college
8	Pre-school	0-5	None	None

From San Diego ABM

- Model segmentation
- Summarize outputs
- Explanatory variables in models
- Constraints on available alternatives

Bridge Expansion Example

- No Build Alternative
  - 4 lanes (2 in each direction, no occupancy restrictions)
  - No tolls
  - Regional transit prices do not change by time of day
- Build Alternative(s)
  - Add 1 lane in each direction (total of 6)
  - New lanes will be HOV (peak period or all day?)
  - Tolling (flat rate or time/congestion-based)
  - Regional transit fares priced higher during peak periods

Bridge Expansion Example—Relevance to Population Synthesis

- Higher income households have a higher willingness to pay—affects distribution of project benefits
- Multiple-worker households are more likely to use HOV
- It is important that a synthetic population adequately represents all relevant sub-groups:
  - Low, medium and high income groups identified at appropriate "breakpoints"
  - The joint distribution of households by age, size, income and workers should match the real population… and by geographic sub-area
  - Household and person attributes should support prediction of relevant variables in other models within the model system:
    - Auto ownership, transit pass holders, telecommuters

What does a synthetic population look like?

Control group levels—may also want (uncontrolled) continuous values for income

HOV Potential?

Household records

TAZ	HHID	Age of Household Head	Number of Persons	Income Group	Presence of Children	Autos Owned	Number of Workers
1433	16670	1	2	1	1	0	1
1937	17392	1	2	2	1	0	1
77	232	1	3	3	1	1	2
18	5042	1	3	2	1	3	2

If finer geographic resolution is desired (e.g., parcels), may need to sub-allocate

Attribute values to be updated by Long-Term models

Use to ID full-time vs. part-time workers

Person records

TAZ	Household ID	Person ID	Age	Works From Home	Employ-ment Status	Gender	Is Student	Hours Worked per Week
77	232	1	22	1	1	2	0	9
77	232	2	24	1	1	1	0	45
77	232	3	1	0	0	2	0	0

Steps in Creating a Synthetic Population

Specify Target Distribution
- control variables
- disaggregate sample
- geographic units

↓

Specify Uncontrolled Variables

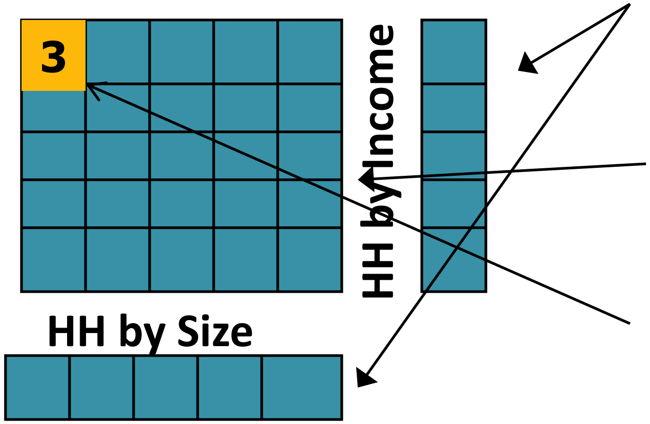
↓

Run Synthetic Population Generator

↓

Modify “Raw” Synthetic Population to Create Final Attribute Variables as Needed

Population Synthesis—Summary View



Step 1 – Begin with **marginal control totals** for each zone and a disaggregate sample (“**seed data**”)

Step 2 – Generate the **joint (multi-way) distribution** of targets for each zone

Step 3 – **Draw** household and person records from disaggregate samples to meet cell quotas

PUMS Household File

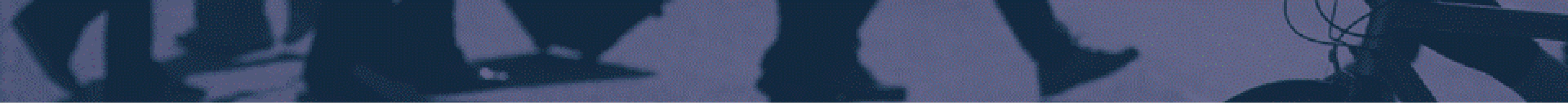
HHID	SIZ	INC	WRK	SF	AGE_HH
1	1	1	1	0	24
2	1	1	0	1	23
3	1	1	0	1	43
4	1	1	1	0	32
5	1	1	1	1	34
6	2	2	2	0	49
7	2	2	2	1	67
8	3	2	2	1	15
9	3	2	2	0	12

Household File

HID	SIZ	INC	WRK	SF	AGE_HH
1	1	1	1	0	24
3	1	1	0	1	43
5	1	1	1	1	34

Person File

HID	PID	AUT	INC	WRK	GEN	AGE	EMP
1	1	1	1	1	0	24	1
3	1	3	1	0	0	43	1
5	1	0	1	1	1	34	0



Steps in Creating a Synthetic Population

Specify Target Distribution
- control variables
- disaggregate sample
- geographic units

↓

Specify Uncontrolled Variables

↓

Run Synthetic Population Generator

↓

Modify “Raw” Synthetic Population to Create Final Attribute Variables as Needed

Specifying a Target Distribution

Identify attributes most important to model design

- What are the important household characteristics?
- What are the important person characteristics?
- Household size, age, workers, and income are typical
- Presence of children, person age and gender desirable

Obtain data needed for control variables and sample

- Controls: Census, ACS, CTPP, other TAZ
- Sample: PUMS, household travel survey
- Geography: GIS equivalencies between spatial units

Identify which attributes to control and at what level

- Household size: 1, 2, 3, 4+ … or .. 5+ ?
- Income level: Lo/Med/Hi? Quartiles? CPI Deflated?
- Presence of children: no/yes … by age group?
- Age by group: 0-4, 5-15, 16-17, 18-24, 25-64, 65+?

Determine level of spatial resolution

- Household size, age of head available at block level
- Income, number of workers available at block group
- Housing units, others available at TAZ

Data Sources for Control Data

- Base year
  - Decennial Census: ~100% sample
  - American Community Survey (ACS) summary files:
    - 3% sample, rolling 5-year sample, yields an estimate of ~15% of pop.
  - Other zonal data developed locally (TAZs)
  - Census Transportation Planning Products (CTPP)
  - TAZ-Census Block/Group geographic equivalencies
- Forecast year
  - Regional socio-economic forecasts, growth factor models, allocation processes
  - Output from a land-use model

Data Sources for Disaggregate Sample

- Census/American Community Survey (ACS) Public Use Microdata Sample (PUMS) data
  - Rolling 3-years (e.g., 2007-2009) or 5-years (2005-2009)
  - Geographically sampled/grouped by public use micro data areas (PUMAs) ~ 100,000 population -/+
- A representative regional household survey could be used for disaggregate sampling
  - Insufficient quantities and sampling biases could be an issue
  - This may be the only source for certain sub-groups
    - e.g., National Household Transportation Survey (NHTS) add-on for Florida includes seasonal households

Specifying Control Attributes(1 of 3)

- Control attributes could be single or multi-dimensional
  - Multi-dimensional attributes can be treated as single dimensional attributes with number of categories equal to the product of the numbers of categories in individual attributes, minus infeasible cells.
  - Advantage: more precise regional control over the correlation between attributes
  - Disadvantages: sparse combinations
    - difficulty in finding samples to meet cell quota
    - may over-represent certain sample households

Size1 Workers0 Income1
Size1 Workers0 Income2
Size1 Workers0 Income3
Size1 Workers0 Income4
Size1 Workers1 Income1
Size1 Workers1 Income2
Size1 Workers1 Income3
Size1 Workers1 Income4
Size2 Workers0 Income1
Size2 Workers0 Income2
Size2 Workers0 Income3
Size2 Workers0 Income4
Size2 Workers1 Income1
Size2 Workers1 Income2
Size2 Workers1 Income3
Size2 Workers1 Income4
Size2 Workers2 Income1

… etc.

Size4 Workers2 Income1
Size4 Workers2 Income2
Size4 Workers2 Income3
Size4 Workers2 Income4

Specifying Control Attributes(2 of 3)

- It is desirable that all control attributes are somewhat “orthogonal” to each other
  - i.e., their variance in the population is largely independent
- Controlling for two attributes that are highly correlated is no better than controlling for just one
  - Example: In a region, if certain income categories are correlated with race, then it may not be efficient to include both income and race of head of household in the control attribute set

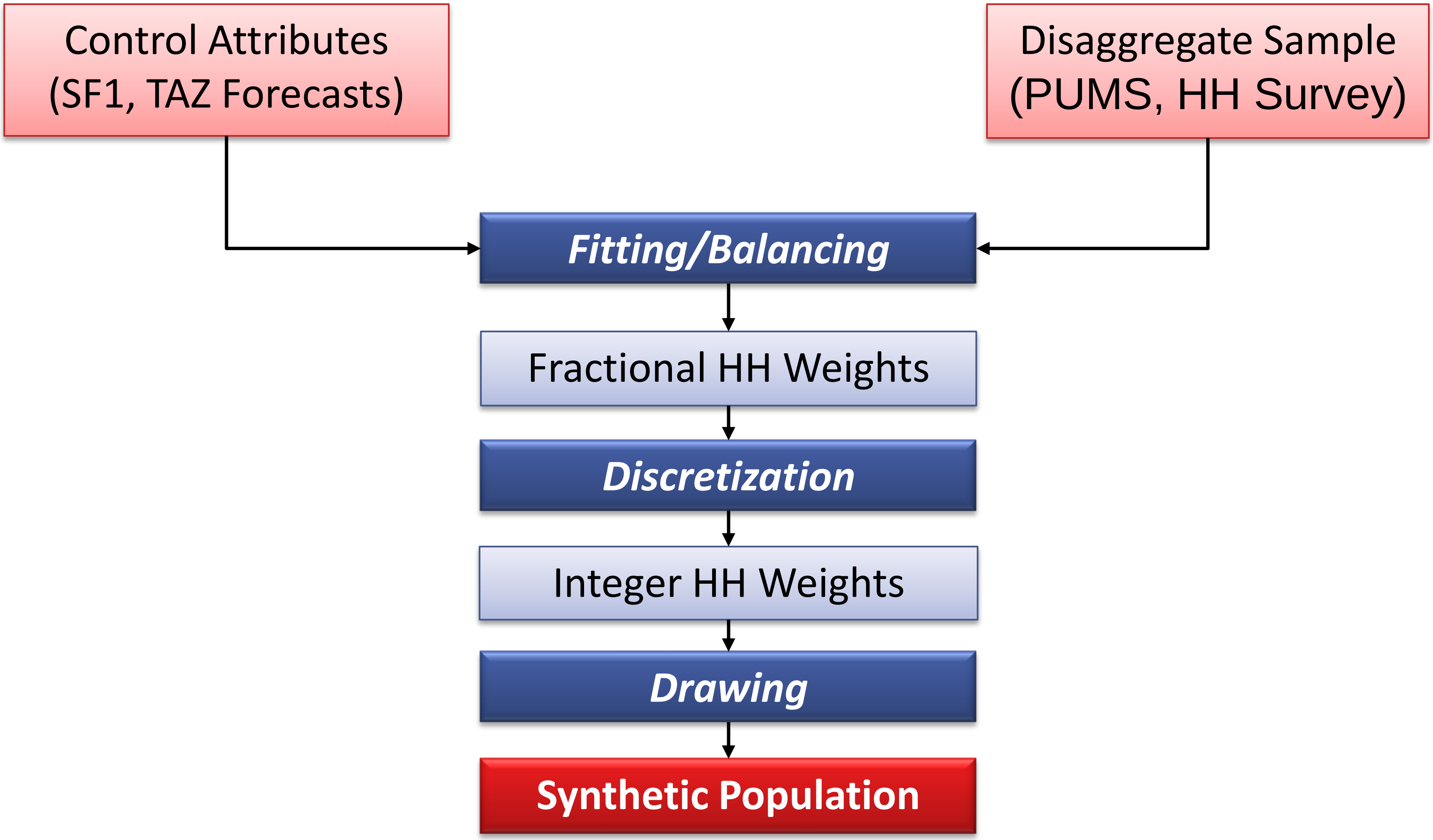
Specifying Control Attributes(3 of 3)

- Number of control attributes
  - Too few control attributes:
    - Synthetic population may not accurately reflect the true population
  - Too many control attributes leads to sparse cells in seed data:
    - May distort true distribution
- Control attributes may use different geographic units
  - If control attribute totals are not accurate at a particular spatial unit, they could be specified at a lower resolution
  - Best if spatial units “nest” (Census blocks & block-groups)
- Different sets of control attributes may be used for base and forecast years, limited by forecasting accuracy

Uncontrolled Attributes

- Uncontrolled attributes are directly transferred to the synthetic population when the sample is drawn
  - All the attributes present in the sample may not be controlled for since it could affect the match negatively
  - Such variables are often needed by activity-based models in addition to the control variables
  - Should be well-correlated with controlled variables
- Examples
  - Person worker status, student status, race, occupation
- Be wary of potential bias due to loose correlation with control variables!

Population Synthesis Process



Iterative Proportional Fitting (IPF)

- Widely used and statistically robust procedure to control the joint distribution of attributes in a synthetic population. (See Beckman et al, 1996)

Basic Steps:

1. Choose a set of control attributes and their levels
2. Obtain control totals for the control attributes by level
3. Obtain a sample of households with the relevant attributes to be the "seed matrix"
4. Create a joint distribution of control attributes from the disaggregate sample data
5. Iteratively factor the cells in the matrix, based on marginal control attribute target values for each attribute level, until all target values are matched or nearly matched

IPF Illustration – Setup

		Household Income			
		High	Low	Total	Household Size Control
Household Size	Adjustment	-	-		
1	-	2	3	5	20
2	-	4	1	5	50
3 or more	-	1	3	4	30
Total		7	7		
Household Income Control		40	60		

Sample

Controls

Controls

IPF Illustration – Row Adjustments

		Household Income			
		High	Low	Total	Household Size Control
Household Size	Adjustment	-	-		
1	20/5 = 4	2	3	5	20
2	50/5=10	4	1	5	50
3 or more	30/4=7.5	1	3	4	30
Total		7	7		
Household Income Control		40	60		

IPF Illustration – Iteration 1

		Household Income			
		High	Low	Total	Household Size Control
Household Size	Adjustment				
1	4	2x4=8	12	20	20
2	10	4x10=40	10	50	50
3 or more	7.5	1x7.5=7.5	22.5	30	30
Total		7	7		
Household Income Control		40	60		

IPF Illustration – Column Adjustments

		Household Income			
		High	Low	Total	Household Size Control
Household Size	Adjustment	40/55.5=0.72	60/44.5=1.35		
1		8	12	20	20
2		40	10	50	50
3 or more		7.5	22.5	30	30
Total		55.5	44.5		
Household Income Control		40	60		

IPF Illustration – Iteration 2

		Household Income			
		High	Low	Total	Household Size Control
Household Size	Adjustment	0.72	1.35		
1		8x0.72=5.8	12x1.35=16.2	22	20
2		28.8	13.5	42.3	50
3 or more		5.4	30.3	35.7	30
Total		40	60		
Household Income Control		40	60		

IPF Illustration – Iteration 13 (Final)

		Household Income			
		High	Low	Total	Household Size Control
Household Size	Adjustment				
1		4.51	15.49	20	20
2		31.70	18.30	50	50
3 or more		3.79	26.21	30	30
Total		40	60		
Household Income Control		40	60		

Drawing Households from a Sample

1. Calculate selection probabilities for each household in the sample
   – Based on the attributes of the household and the number of such households in the joint distribution
2. Draw households from a sample
   – Based on the selection probabilities to match target numbers by each household type in the joint distribution
   – e.g., Monte-Carlo process, cycling methods
3. Output a synthetic population consisting of all persons belonging to the households drawn

Methods for Estimating a Joint Distribution

- Matrix balancing
  - Follows the standard IPF approach
  - Involves creation of multidimensional matrices during fitting
  - Controls either household or person attributes, but not both
- List balancing
  - Involves applying IPF on individual households in a list
  - More complex algorithm
  - Both household and person attributes can be simultaneously controlled
  - Reduces sparse matrix problems

Other Ways to Sample from a Distribution

- **Intelligent drawing**
  - Directly draw households from the sample to match control totals, but without a separate fitting/balancing step (no IPF)
  - Drawing informed by the extent to which a sample household would contribute to the match of both household and person controlled attributes
    - Examples: ARC (Atlanta), FSUTMS (Florida)

Other Ways to Sample from a Distribution

- **Discretizing**
  - Developing individual household weights which are then "discretized" so that the distributions with respect to controlled attributes are preserved
    - (e.g. "intelligent bucket rounding")
  - Individual weights can be developed by IPF or entropy maximization techniques
  - Similar to list balancing, but does not involve a drawing process
    - Example: SANDAG

Questions and Answers

The Travel Model Improvement Program

Implementations of Population Synthesizers

Type of Synthesizer	Fitting and Drawing	Fitting only	Drawing only
Matrix & table balancing	• PopSynWin (U. Illinois-Chicago) • CEMSELTS (U. Texas-Austin) • TRANSIMS (FHWA) • ALBATROSS (TU-Eindhoven) • MORPC (PB)		• ARC PopSyn (PB)
List balancing	• PopGen (Arizona St.) • ILUTE (U. Toronto)	• SANDAG PopSyn (PB)	• FSUTMS (U. Florida)

Balancing Procedures

Generalization

Type	Controls	A priori weights (seed distribution)	Contribution coefficients
Multidimensional Matrix (MORPC)	Row/column totals	Initial matrix	Cell-row/column incidence (0,1)
Table of categories (ARC)	Column totals	Initial weight for category (row)	Row/column incidence (0,1)
List of individual records (SANDAG)	Column totals	Initial individual weight (row)	Row/column coefficient (≥0)

- Each subsequent method includes the previous one as a particular case and guarantees the same result
- Not every table of categories can be reduced to a matrix form!
- Not every table of individual records can be reduced to table of categories!

Example of List of Individuals for Balancing

HH ID	HH size				Person age				HH initial weight
	1	2	3	4+	0-15	16-35	36-64	65+	
	$i=1$	$i=2$	$i=3$	$i=4$	$i=5$	$i=6$	$i=7$	$i=8$	ω_n
$n=1$	1							1	20
$n=2$		1			1	1			20
$n=3$			1			1	2		20
$n=4$				1		2	2		20
$n=5$				1	1	3	2		20
….									…
Control	100	200	250	300	400	400	650	250	

Program Formulation (Fixed Controls)

$$\min_{\{x_n\}} \sum_n x_n \ln \frac{x_n}{w_n}$$ - Preserve initial weights as much as possible

$$\sum_n a_n^i x_n = A^i, (\alpha^i)$$ - Meet all controls

$$x_n \geq 0$$

- Convex mathematical program with linear constraints
- Solution can be found by forming the Lagrangian and equating partial derivatives to zero (necessary conditions)
- Conventional matrix balancing or table balancing are particular cases

Advantages of Balancing List of HHs

- No reason to fight zero cells in joint distributions, they cannot be utilized anyway
- Can incorporate HH-level and person-level controls naturally
- Prepares background for discretizing; no need in drawing

Relaxation of Controls (MAG, BMC)

- Objective function: $$min \sum_n x_n ln \frac{x_n}{w_n} + \sum_i \alpha_i Y_i ln Y_i$$
- Match relaxed controls: $$\sum_n a_{ni} x_n = A_i Y_i, (\lambda_i)$$
- HH weights and relaxation factors: $x_n \geq 0,\quad Y_i \geq 0$
- Importance factors for controls: $\alpha_i \geq 0$
  - Set to a large value of 1,000 to ensure match if feasible
  - Set to 1,000,000 for total number of HHs

Importance of Entropy-Maximizing Balancing vs. Simple IPF

Household type	Initial distribution	Linear IPF	Entropy-max balancing
0 workers	20%	4%	12%
1 worker	30%	36%	25%
2 workers	40%	48%	47%
3 workers	10%	12%	16%
Control average	1.4	1.8	1.8

- That's how entropy-maximization works!

Implementation Example 1:
Baltimore Metropolitan Council (BMC)

- Implemented using PopGen (ASU) at the TAZ level
- Only household attributes controlled
  - Household size (5 categories)
  - Income (4 categories)
  - Workers (4 categories)
- Joint totals of size-income and income-workers from the synthetic population used in trip generation models (4-step)

Implementation Example 1:

Baltimore Metropolitan Council (BMC)

Variable	Control	Synthetic	% Difference
Household Size			
1 person	527,210	527,266	0.01%
2 persons	561,788	562,293	0.09%
3 persons	333,499	333,607	0.03%
4 persons	261,710	261,534	-0.07%
5 or more persons	207,021	206,528	-0.24%
Total	1,891,228	1,891,228	0.00%
Household Income			
< $11,800	190,133	189,005	-0.59%
≥ $11,800 and < $26,000	284,592	283,968	-0.22%
≥ $26,000 and < $44,200	378,935	378,845	-0.02%
≥ $44,200	1,037,528	1,039,410	0.18%
Total	1,891,188	1,891,228	0.00%

Implementation Example 2:
Atlanta Regional Council (ARC)

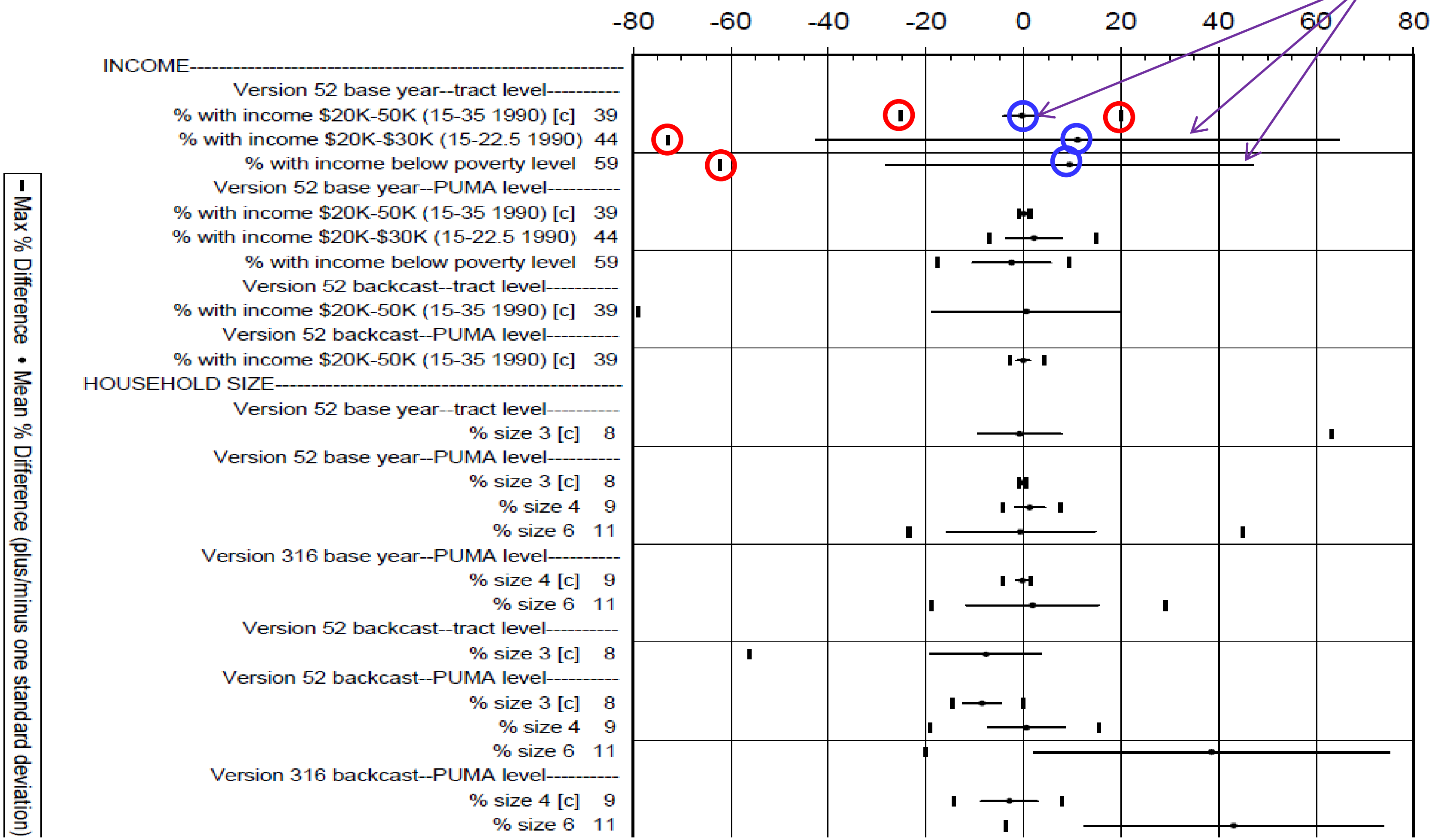
- Implemented using PopSyn at the TAZ level
- Household attributes controlled
  - Household size (5 categories)
  - Income (4 categories)
  - Workers (4 categories)
  - Child presence (2 categories)
  - Age of head of household (2 categories)
  - Family/Non-family (2 categories)

Max % diff

Min % diff

Implementation Example 2:
Atlanta Regional Council (ARC)

Standard Dev.



Implementation Example 2:
Atlanta Regional Council (ARC)

- Lessons learned from "backcasting"
  - Bias arising from uncontrolled variables
    - Race was uncontrolled and backcasts revealed gross inaccuracies in projections
  - Bias arising from more aggregate controls for forecast year
    - The seed preserves the distribution from the base year and the distribution of the uncontrolled variable changes in the future year within the more aggregate category
  - Practical implications:
    - If uncontrolled variables are not directly used in the model these biases are tolerable
    - Uncontrolled variable with a strong impact on travel demand should be better re-specified and additional controls introduced

Implementation Example 3:
SHRP 2 C10A – Jacksonville, Florida

- Implemented using PopGen at the TAZ level
- Includes seasonal households explicitly:
  - Have very different HH structure & travel behavior
  - Relevant for certain seasons only
- Household attributes controlled
  - Household size-income-workers (44 categories)
  - Child presence (2 categories)
  - Age of head of household (3 categories)
- Person attributes controlled
  - Age (5 categories)
  - Gender (2 categories)

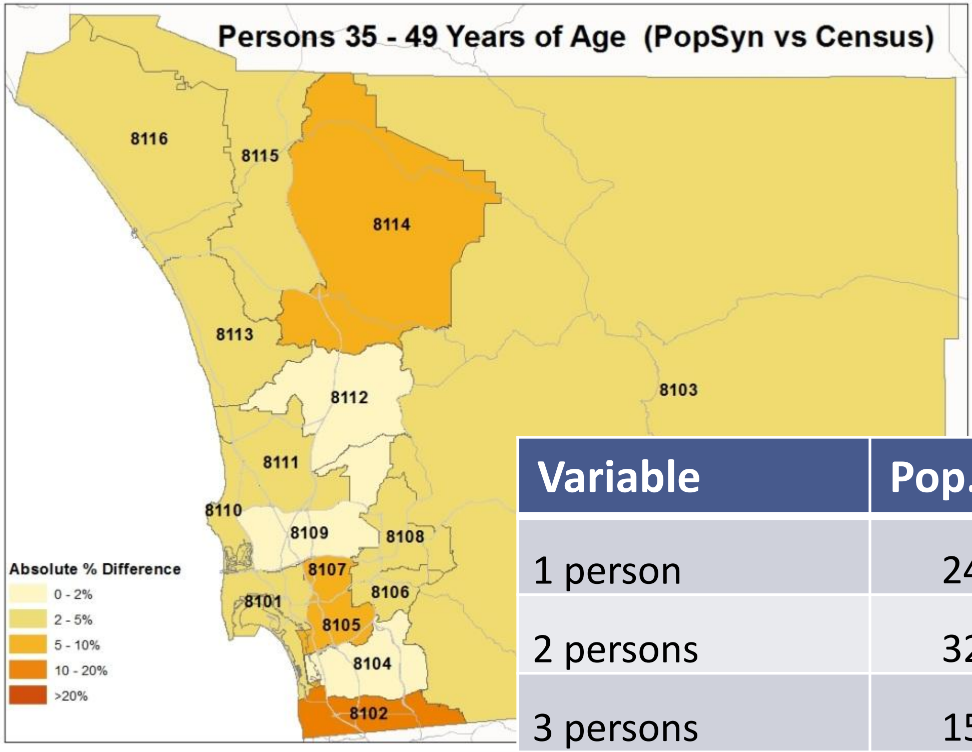
Implementation Example 3:
SHRP 2 C10A – Jacksonville, Florida

Variable	Control	Synthetic	% Difference
Household Size			
1 person	118,841	119,901	0.89%
2 persons	161,113	161,595	0.30%
3 persons	84,235	83,394	-1.00%
4 or more persons	115,067	114,408	-0.57%
Total	479,255	479,298	0.01%
Person Age			
0-15 years	286,068	283,248	-0.99%
16-20 years	78,668	77,511	-1.47%
21-44 years	443,351	435,734	-1.72%
45-64 years	270,899	266,070	-1.78%
65+ years	123,868	122,237	-1.32%
Total	1,202,855	1,184,800	-1.50%

Implementation Example 4:
San Diego Assoc of Governments (SANDAG)

- Implemented using PopSyn II at the TAZ level
- Household attributes controlled
  - Household size (4 categories)
  - Income (5 categories)
  - Workers (4 categories)
  - Child presence (2 categories)
  - Dwelling unit type (3 categories)
  - Group quarter status (4 categories)
- Person attributes controlled
  - Age (7 categories)
  - Gender (2 categories)
  - Race (8 categories)

Implementation Example 4:
San Diego Assoc of Governments (SANDAG)



Variable	Pop. Syn.	Census	Mean diff.	Std. Dev.
1 person	24.2%	24.2%	-0.4%	1.5%
2 persons	32.3%	32.0%	0.8%	1.0%
3 persons	15.9%	16.1%	-1.8%	2.0%
4 or more persons	27.7%	27.7%	-0.7%	3.3%
Total HHs	985,938	992,681	-0.6%	0.9%

Why Good Fit is Important

- The outputs of a synthetic population are the inputs to all other model components.
  - If these inputs are inaccurate, forecasts and other analyses will also be inaccurate.
- Types of Errors:
  - Under-representation of market sub-segments → model may be insensitive
  - Over-representation of market sub-segments → model may be too sensitive, or sensitive in unexpected ways
  - Mis-alignment of population with geography → inaccurate forecasts (trip lengths, mode shares, etc.)

Getting good fit… practical tips!

- Choosing appropriate control attributes
  - Controlling for certain attributes may distort the distributions of others
  - Several iterations of testing may be required before determining the final set of control attributes
  - Meta-analysis of consistency between the controls
  - Differential importance weights can be applied to controls
- Zero cells in the seed matrix created from sample data
  - Redefine or combine attribute categories (collapse attribute levels to fewer groupings)
  - Expand the geographic unit to a more aggregate level

Getting good fit… more practical tips

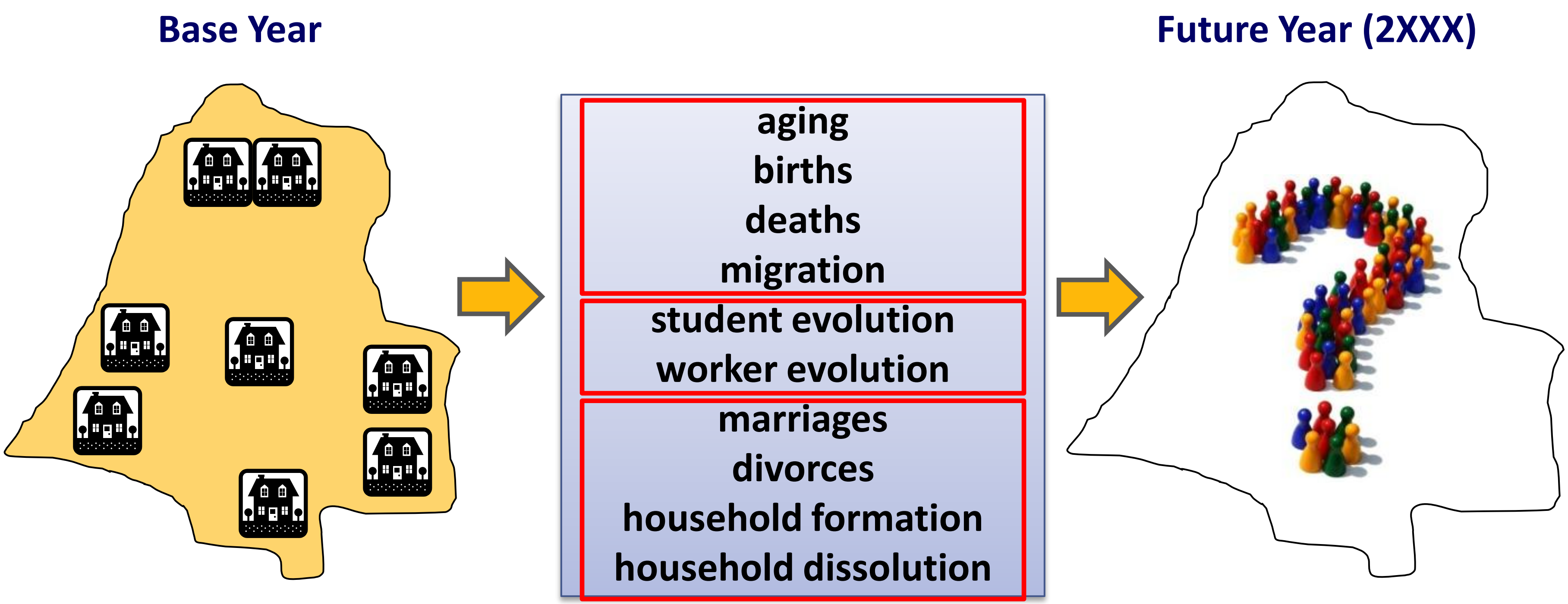
- It is usually possible to maintain tight control on either household control targets or person control targets, but not both simultaneously:
  - Importance weights higher for main HH controls
  - Importance weights lower for secondary HH controls and person controls
- It is possible to indirectly control person attributes at household level
  - In practice, the only person characteristics that are usually controlled for are age and gender.
    - Gender should work out without explicit controls.
    - Age can be handled at the household level by using age of head and presence/absence of people in certain age groups (i.e., very young children, school age children, etc.)

Forecasting Future Synthetic Populations

- Where do we get marginal control totals and sample populations for future years?
  - Control Variables:
    - Less controls are normally set for future years
    - Trend extrapolation based on horizon-year forecasts for entire region (or sub-region if available)
    - Land Use and Demographic model outputs
  - Sample population:
    - Re-use… assume similar joint distributions of household and person attributes in the horizon year… appropriate in built-out areas
    - Consider changing demographics… enrich the sample by adding HHs from other geographic units that look more like your region in the future (primarily, ethnicity or income mix)
  - Household evolution models (emerging method)

Household Evolution

- Synthetic population for a base year is evolved to a future year by simulating certain demographic evolutionary processes



Why Household Evolution?

- Demographic shifts, such as the aging of the population, are not always reflected in non-evolution forecasting processes
- Create new HHs by logical combination of features instead of locking in a relatively small sample
- Most models fail to account for in- and out-migration
- Socio-economic forecasts developed more methodically than applying naive growth rates to marginal distributions
- Potential to provide more variables for dynamic travel models – variables of change and lagged variables
- Consistent when integrated with land use models, which are also usually evolution models

Cohort Survival Method

- Standard demographic projection method
  - **Aging, fertility, mortality** and **net migration** by age cohort

population at the beginning of the time period

population at the end of the time period

cohort i	Age group	Pop t_0	Survival Rate	Survive to t_{0+10}	Birth Rate	Births	Net Migrati	Pop t_{0+10}
1	0 - 9	3,900	0.989		0	0	5	425
2	10 - 19	3,200	0.999	3857	0.011	35	0	3,857
3	20 - 29	3,300	0.998	3197	0.081	267	50	3,247
4	30 - 39	2,800	0.998	3293	0.038	106	35	3,328
5	40 - 49	1,700	0.996	2794	0.007	12	10	2,804
6	50 - 59	1,800	0.991	1693	0	0	0	1,693
7	60 - 69	1,100	0.975	1784	0	0	-20	1,764
8	70 - 79	550	0.936	1073	0	0	0	1,073
9	80+	200	0.88	691	0	0	0	691
	TOTAL	18,550		18,382		420	80	18,882

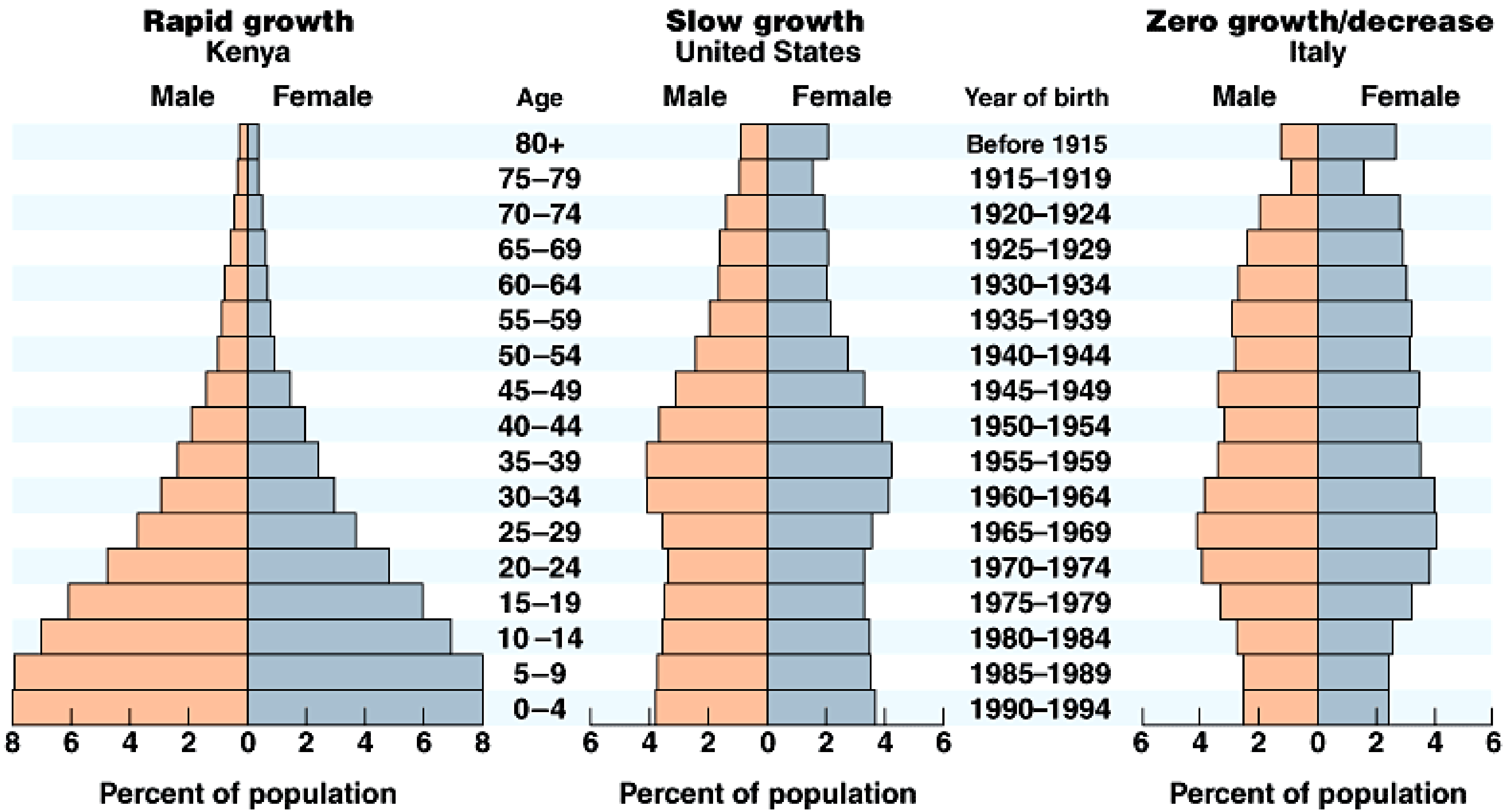
to calculate births:

births by parent age cohort group:
10-19: 3200 * .011 = 35
20-29: 3300 * .081 = 267
30 -39: 2800 * .038 = 106
40-49: 1700 * .007 = 12
the sum of these is <u>420</u>

the population at the end of the time period then becomes the population at the beginning of the next time period

Source:
http://www-personal.umich.edu/~sdcamp/up504/cohort%20survivalexamplew08.pdf

Population Pyramids



How to reflect evolutionary processes in a future synthetic population?

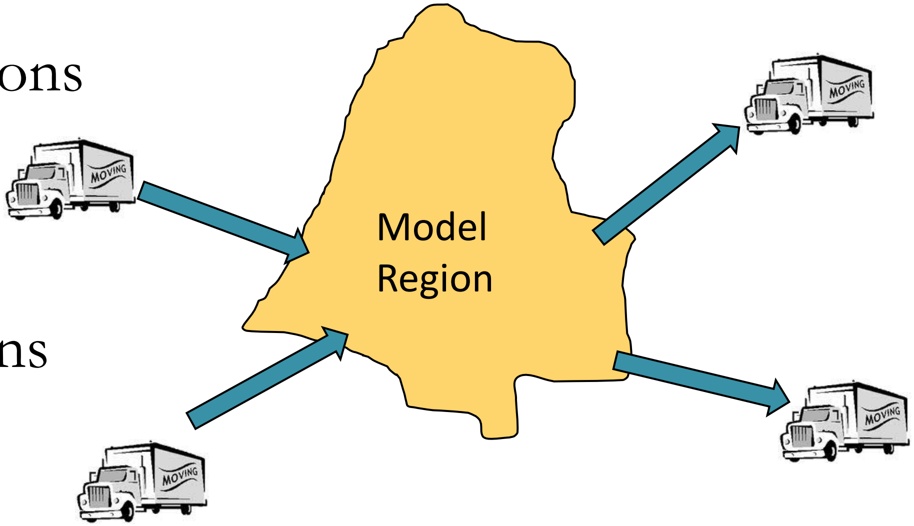
- Use cohort survival with net migration to forecast future controls for person-level variables
- What about households?
  - First attempts to model household formation (marriage, divorce, children, etc) explicitly
- More information needed on household formation and dissolution
  - Correlate trends in person evolution with evolutionary processes in households

Person Evolution

- Aging and Mortality
  - Aging simply adds one to the age of the population each year
  - Mortality rate-based probability or survival model
  - Challenge – changing mortality rates due to medical advances
- Birth
  - Determines if a female gives birth in a given year
  - Challenges:
    - Persons delaying parenthood (socio-economic and cultural tendencies)
    - Incorporating children born outside of marriage/cohabitation (adoptions, foster parents, etc.)

Emigration and Immigration

- Represents migration of population into and out of the model region
- Rate-based probability models
- Challenges:
  - uncertain economic conditions
  - attributes of emigrants
  - attributes of immigrants
  - disaggregation to sub-regions



Student Evolution

- Education
  - Determines education level of a person in the population
  - Challenge – representing people who discontinue education temporarily and later return to school
- College Students' Residential Choice
  - Determines whether children starting college stay at home or move into a dormitory/rent apartment/new household:
    - MAG Population Synthesizer under development
    - Essential for modeling major universities that generate significant student population around the campus

Worker Evolution

- Labor Participation
  - Predicts whether a person joins the labor force
- Occupation Choice
  - Determines the occupational category of a person who is predicted to join the labor force
- Employee Income
  - This model predicts the income/eraning level of an employee which can be used to estimate the household income

Household Formation and Dissolution

- Family formation/Cohabitation/Marriage
  - Determines marriage/cohabitation decision of adults in the population
  - Roommates/non-family households
- Divorce/Household dissolution
  - Determines divorce decision of adults in the population

Household Evolution--Other Challenges

- Choice of model form – discrete choice, regression, hazard-based
- Model sequencing and frequency of simulation
- Validation of the framework and models would itself pose a challenge

Review: Learning Outcomes

- What a synthetic population is and how they are used in activity-based models
- The methods used to synthesize a population and the various considerations with respect to specifying attributes, including:
  – Specifying controlled and uncontrolled attributes
  – Spatial unit of analysis considerations
  – Methods used in synthetic population generators
  – Challenges to good fit
- Emerging methods in household evolution and why they might be important

Questions and Answers

The Travel Model Improvement Program

2012 Activity-Based Modeling Webinar Series

Executive and Management Sessions	
Executive Perspective	February 2
Institutional Topics for Managers	February 23
Technical Issues for Managers	March 15
Technical Sessions	
Activity-Based Model Frameworks and Techniques	April 5
Population Synthesis and Household Evolution	April 26
Accessibility and Treatment of Space	May 17
Long-Term and Medium Term Mobility Models	June 7
Activity Pattern Generation	June 28
Scheduling and Time of Day Choice	July 19
Tour and Trip Mode, Intermediate Stop Location	August 9
Network Integration	August 30
Forecasting, Performance Measures and Software	September 20

Continue the discussion online…

The new TMIP Online Community of Practice includes a Discussion Forum where members can post messages, create forums and communicate directly with other members. Simply sign-up as a new member, navigate to http://tmiponline.org/Community/Discussion-Forums.aspx?g=posts&t=523 and begin interacting with other participants from today's webinar session on Activity-Based Modeling.