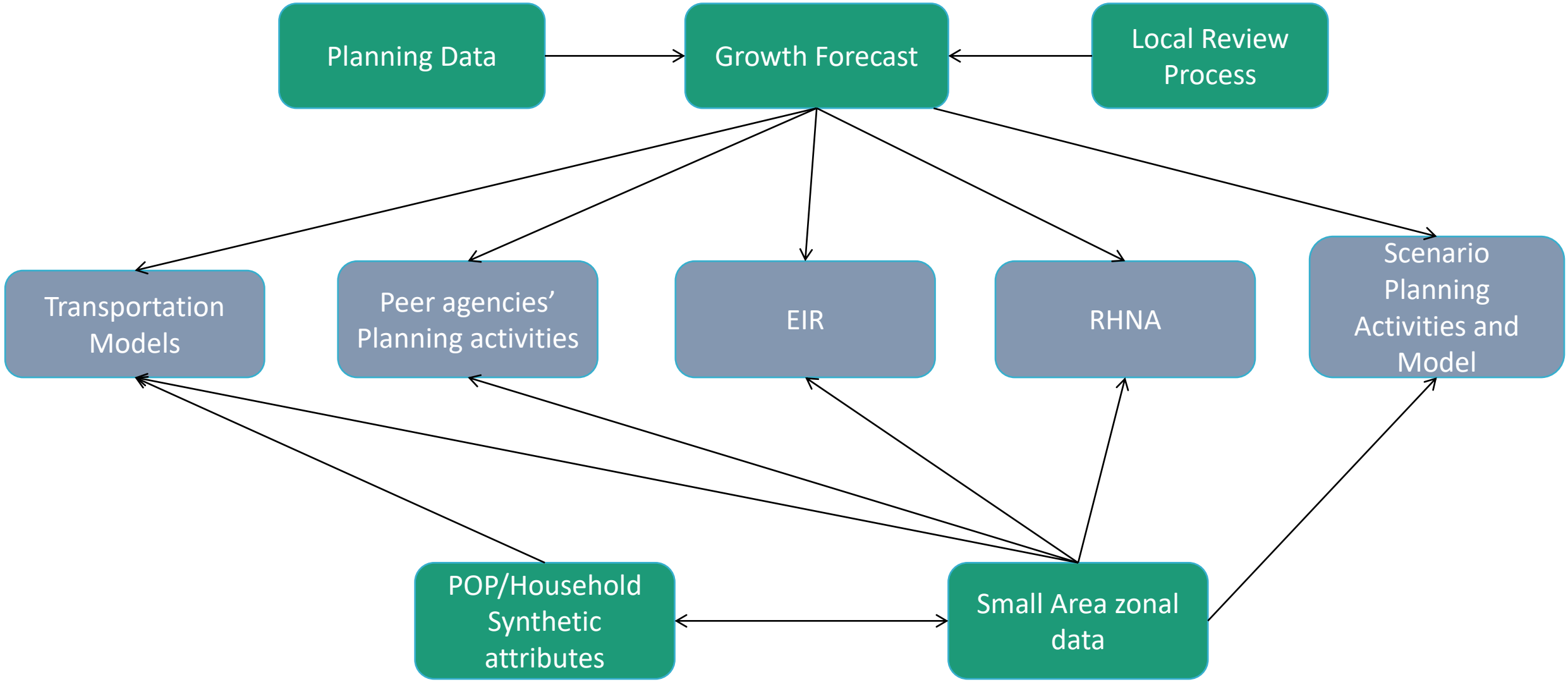
SCAG SOCIOECONOMIC DATA DEVELOPMENT FOR TRANSPORTATION MODEL

Modeling Task Force Meeting
September 28, 2022

Forecasting Unit
Modeling & Forecasting
SCAG

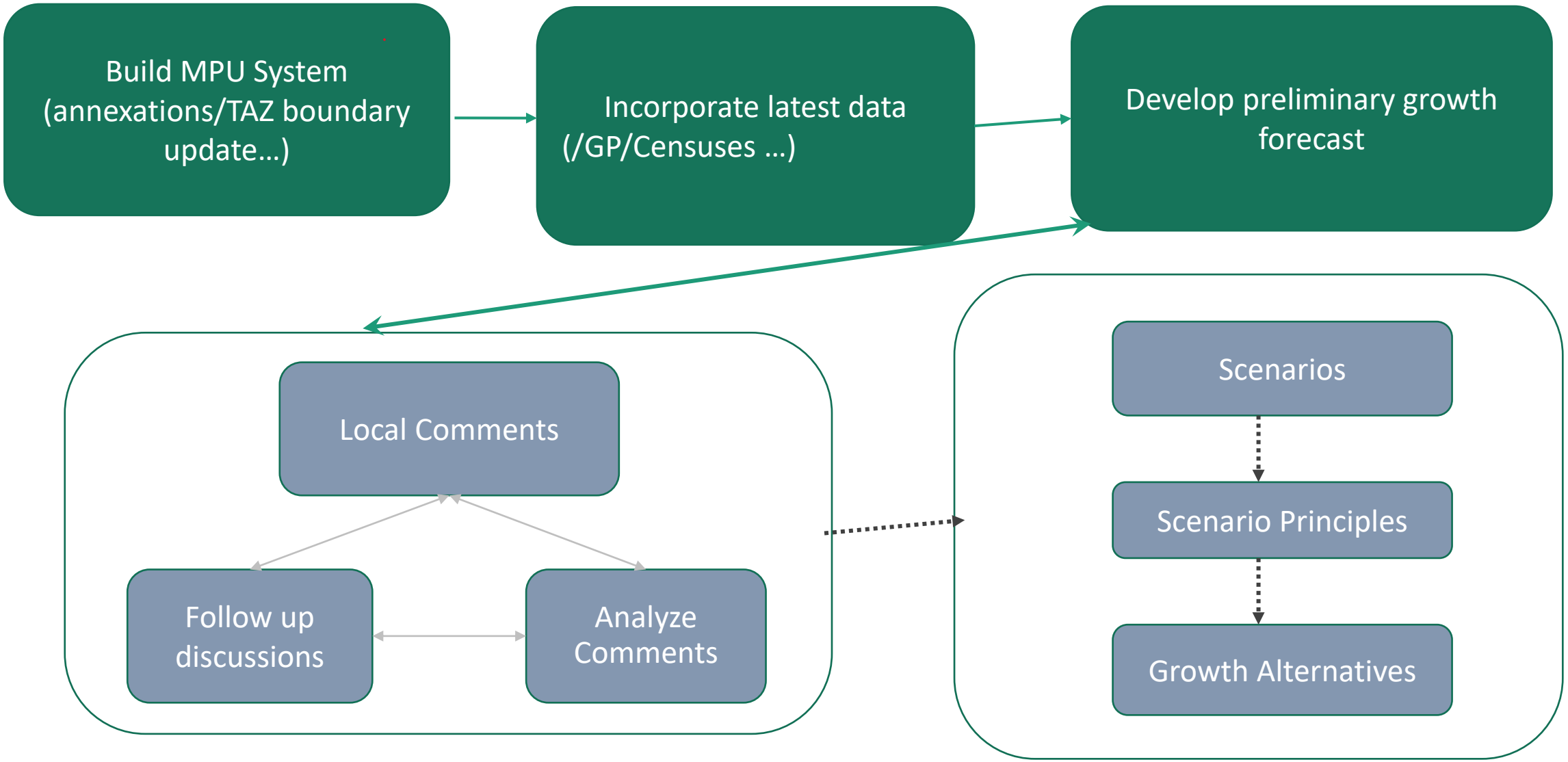
Growth Forecasting Functions



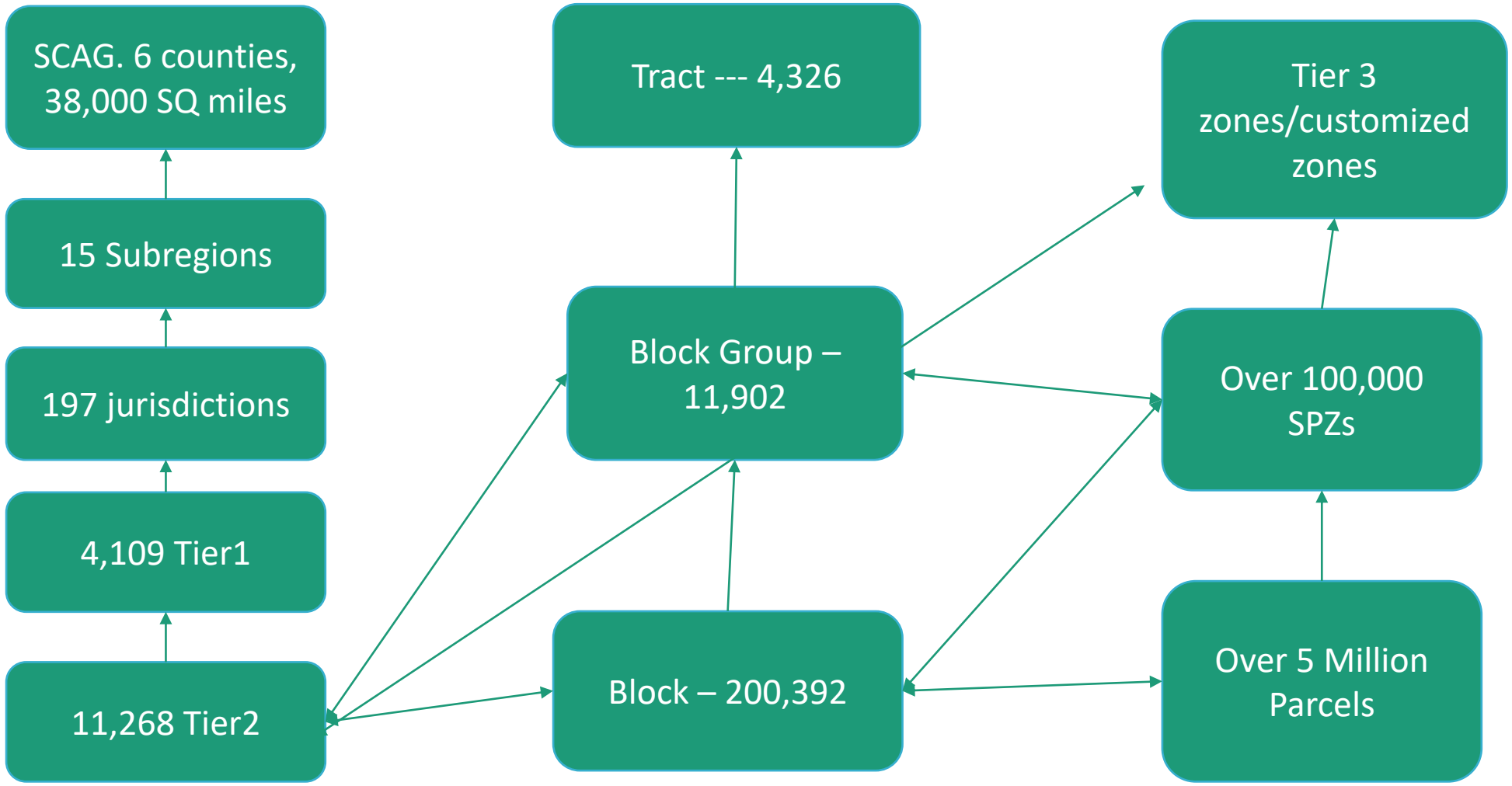
Sub-County Growth Forecast Goals

- Under CEHD, TWG, and other SCAG committees' guidance, develop Sub-county levels including jurisdictions and TAZ socioeconomic estimates and projections
- Build growth projection analytical framework and database for federal and state mandated long-range planning
  - ✓ Regional Transportation Plan /Sustainable Communities Strategy (RTP/SCS)
  - ✓ Air Quality Management Plan (AQMP)
  - ✓the Federal Transportation Improvement Program (FTIP)
  - ✓the Regional Housing Needs Assessment (RHNA)

Sub-County Growth Forecast Development



SCAG Region Geographical Zones



County Level Projections

Input

Key Controls: POP, Household, EMP
(Expert panel/local input)

DOF: Population Projection, 2010-2060 (by Age, Gender, and Race/Ethnicity)

ACS PUMS: Distributions by Age, Gender, and Race/Ethnicity

Decennial Census: Demographic correlation distributions

Projection

POP:
Age, Race/Ethnicity/Worker status, Group Quarter

Household:
Size, Type, Income, age

Worker by Industries

Socioeconomic Data Process

Input

County Level Projections

- Pop by age, race/ethnicity
- Household by type
- EMP by sectors
- Residential Pop, GI, GN
- K12, College Enrollment
- Other secondary variables

Tier2 Level Projections

- Pop, Household, EMP, Sectors
- Residential Pop, GI, GN
- K12, College Enrollment

Census Decennial Data

Census block level demographic interrelation

ACS Data

- PUMS
- Current 5-year data

PopSyn Control (Tier2)

Probabilistic Choice Models (Multinomial Logit Model)

POP: Age, Race/Ethnicity, Gender

Household: Size, Housing Type, Income, Age

Worker: Worker status, Earning

Employment: Earning

Constant Share Method

Group Quarters in Dormitories

PopSyn Output (ABM / TBM)

POP:
Age, Race/Ethnicity,
Group Quarters in Dormitories

Household:
Size, Housing Type, Income, Age
# of Workers, Children, College Students

Worker:
Worker status, Earning

Employment:
Earning, Sectors

Enrollment: K12 and College

TAZ level Growth Forecast

- ✓TBM
- ✓Environment Justice(EJ)
- ✓AQMP
- ✓ABM
  - --- POPsyn Process

PopSyn, what for?

- Creates data of individual households and population
  - Similar to the Household Travel Survey (excluding travel-related questions)
  - Number of records of household file is same to the number of estimated (or projected) household
  - Consistent to the tables summarized by geography (like Census Tract, TAZ)

- Inputs to the ABM
- Reaggregated into summary tables, input to the TBM

- “PopSyn” for Population Synthesis.
  - But, it produces synthetic households
  - By sampling, not by assembling
  - Name is relic from former generation, “PopGen”

What is inside?

- Technical Aspects…
  - Written in Python 3.8 with numpy / pandas.
    - No dependency to DB
  - Control targets at multiple geographies
  - Distribution Synthesis results to smaller geography
  - Discretize (integerize) the household weights

- Methods IPF/IPU/LB (List Balancing)

Multi-geography Control Targets

- Region-Level:
  - Households by number of workers
  - Workers by 20 industrial sectors,
  - Especially workers reside in Imperial County, in Version 3.8
- County-Level:
  - Median household income
- TAZ-Level:
  - Households by size (5), income (5), housing type (4)
  - Persons by age group (5), race/ethnicity (6), worker

Simultaneous LB

- Collect all controls and do LB at PUMA
  - Region/County-Level controls to PUMA, and TAZ
  - TAZs within a PUMA share the PUMS 5% sample
  - Synthesis of PUMA is not same to the sum of TAZ Synthesis

- Way to reconcile the discrepancy
  - Version 3: Run LB for all TAZs excepting the largest one (TAZ with most household).
  - Version 3.8: Run LB for all TAZs simultaneously, adjust the weights to maintain the PUMA level synthesis over iteration.

Discretize the Household Weights

- Weight of household sample is the product of LB. Weight is real number with fraction
- Version 3
  - A linear optimization module to solve a penalty function to get [0, 1]
  - Caused system halt in some cases.
- Version 3.8
  - A combinatorial optimization module generates “many” sets of [0, 1] and evaluate the set against a penalty function.

IPF / IPU / List Balancing

- IPF (Iterative Proportional Fitting) to adjust known distribution according to new marginal totals.
  - OD matrix based on distance matrix, according to Origin / Destination
  - Joint table of households by [household income] and [household size], based on 2016 PUMS distribution, for 2050 projection.

- IPU (Iterative Proportional Update) to adjust weight of samples to meet the marginal controls proportionally
  - Adjust the weights by control variables of households and population sequentially

IPF / IPU / List Balancing

- Illustration of IPU Algorithm

Variable	Weight	Household Type 1	Household Type 2	Person Type 1	Person Type 2	Person Type 3	Weight 1	Weight 2	Weight 3	Weight 4	Weight 5	Final Weight
Household ID												
1	1	1	0	1	1	1	11.67	11.67	9.51	8.05	12.37	1.36
2	1	1	0	1	0	1	11.67	11.67	9.51	9.51	14.61	25.66
3	1	1	0	2	1	0	11.67	11.67	9.51	8.05	8.05	7.98
4	1	0	1	1	0	2	1.00	13.00	10.59	10.59	16.28	27.79
5	1	0	1	0	2	1	1.00	13.00	13.00	11.00	16.91	18.45
6	1	0	1	1	1	0	1.00	13.00	10.59	8.97	8.97	8.64
7	1	0	1	2	1	2	1.00	13.00	10.59	8.97	13.78	1.47
8	1	0	1	1	1	0	1.00	13.00	10.59	8.97	8.97	8.64
Output measure												
Weighted sum		3.00	5.00	9.00	7.00	7.00						
Constraints		35.00	65.00	91.00	65.00	104.00						
δ_b^a		0.9143	0.9231	0.9011	0.8923	0.9327						
Weighted Sum 1		**35.00**	5.00	51.67	28.33	28.33						
Weighted Sum 2		35.00	**65.00**	111.67	88.33	88.33						
Weighted Sum 3		28.52	55.38	**91.00**	76.80	74.39						
Weighted Sum 4		25.60	48.50	80.11	**65.00**	67.68						
Weighted Sum 5		35.02	64.90	104.84	85.94	**104.00**						
δ_a^b		0.0006	0.0015	0.1521	0.3222	0.0000						
Final weighted sum		35.00	65.00	91.00	65.00	104.00						

NOTE: δ = deviation measure. Text in bold signifies that the weighted sum for a control variable has been matched with the corresponding constraint.
[a]Average of $\delta_b = 0.9127$.
[b]Average of $\delta_a = 0.0954$.

Ye, X., K. C. Konduri, R. M. Pendyala, B. Sana, and P. Waddell. Methodology to Match Distributions of Both Household and Person Attributes in Generation of Synthetic Populations. Presented at 88th Annual Meeting of the Transportation Research Board, Washington, D.C., 2009

IPF / IPU / List Balancing

- List Balancing
  - Things similar to IPU
    - Recreates estimate (synthesized variables), compares to controls, updates the weights
    - Sequential updates by each control variables
  - Things not same to IPU
    - The updating factor is calculated according to the solution of “Entropy” optimization problem.

$$X = \sum_i w_i \cdot a_{ij} \ , Y = \sum_i w_i \cdot a_{ij}^2$$

$$\exp(\lambda_j) = 1 - \frac{\sum_i w_i \cdot a_{ij} - b_j \cdot \gamma_j}{\sum_i w_i \cdot a_{ij}^2 + b_j \cdot \gamma_j \cdot \frac{1}{\alpha_j}}$$

$$= 1 - \frac{X - b_j \cdot \gamma_j}{Y + b_j \cdot \gamma_j \cdot \frac{1}{\alpha_j}} = c_j$$

$$\hat{w}_i = w_i \cdot c_j^{a_{ij}} = w_i \cdot \exp(\lambda_j \cdot a_{ij})$$

$$\hat{\gamma}_j = \gamma_j \cdot c_j^{-\frac{1}{\alpha_j}} = \gamma_j \cdot \exp(\lambda_j \cdot -1/\alpha_j)$$

Thank YOU!

Ying Zhou, zhou@scag.ca.gov, 213/236-1943

Sungbin Cho, cho@scag.ca.gov, 213/236-1989