7th International Conference on Innovations in Travel Modeling (ITM)

June 26, 2018, Atlanta, GA

Using a Population Synthesizer to Derive Robust and Consistent Weights for Survey Sample Data

Presented by:

Daehyun You, *Maricopa Association of Governments*
Ram M. Pendyala & Taehooie Kim, *Arizona State University*
Anurag Komanduri, *Cambridge Systematics, Inc.*
Karthik C. Konduri, *University of Connecticut*

Project Team:

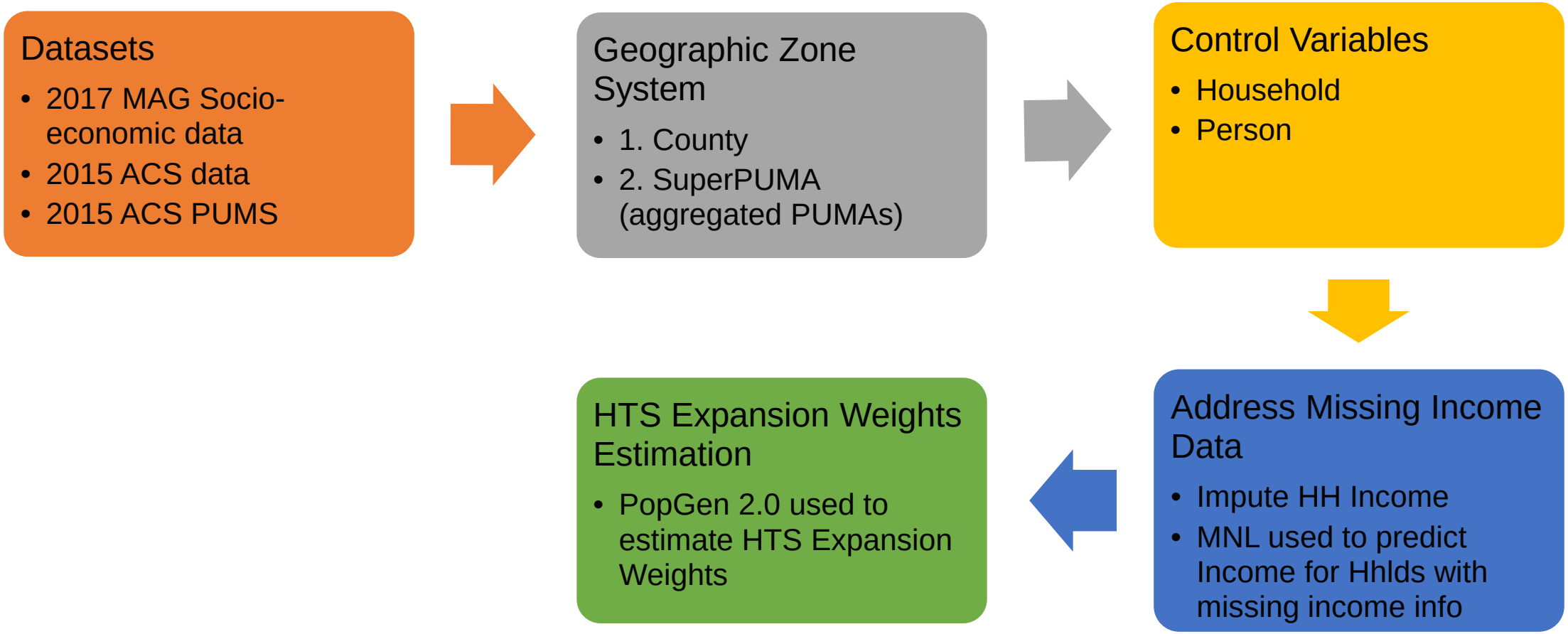
Acknowledgements

- Aman Srivastava, Maricopa Association of Governments
- Edward Brown, Maricopa Association of Governments
- Petya Maneva, Maricopa Association of Governments

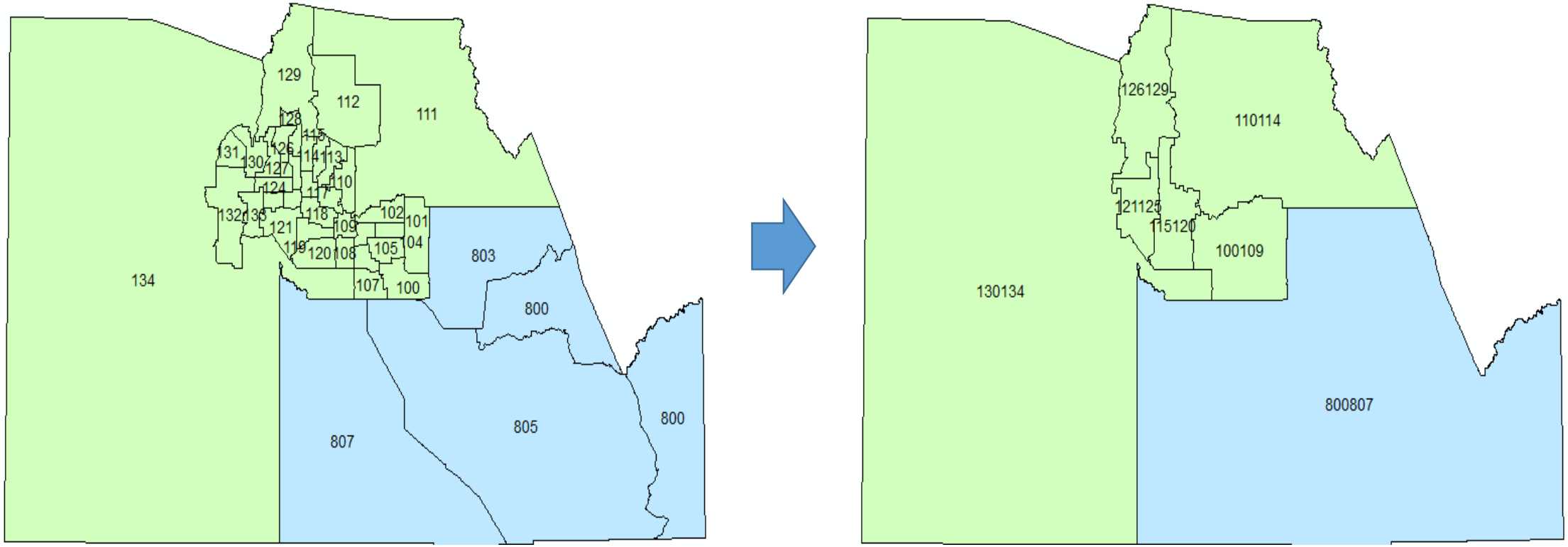
MAG Household Travel Survey (MAG-HTS)

- Maricopa Association of Governments (MAG) conducted a household travel survey in 2016-2017
- GPS activity-travel data collected for two or more days using smartphone app coupled with online activity-travel validation
- Survey data includes:
  - 6,073 complete households
  - 15,097 persons residing in complete households
  - 76,743 stops reported by persons
- As with any survey sample, weighting and expansion is done to reflect population characteristics

Procedure for Computing MAG-HTS Expansion Weights



Geographic Zone System for Expansion Weights



PUMA

SuperPUMA

Control Variables

Household Variables

Geographical Level	Variable Name	Category
SuperPUMA	Household Income	1 = Low (< $35K), 2 = Medium ($35K - <$75K), 3 = High (≥ $75K)
County	Child Presence in Household	1 = No, 2 = Yes
	Household Race (Maricopa)	1 = White & Hispanic, 2 = White & Not Hispanic, 3 = Non-White & Hispanic, 4 = Other
	Household Race (Pinal)	1 = White & Hispanic, 2 = White & Not Hispanic, 3 = Other
	Household Size	1 = 1 person, 2 = 2 persons, 3 = 3 persons, 4 = 4 or more persons
	# Vehicles	1 = 1 vehicle or less in household, 2 = 2 or more vehicles in household

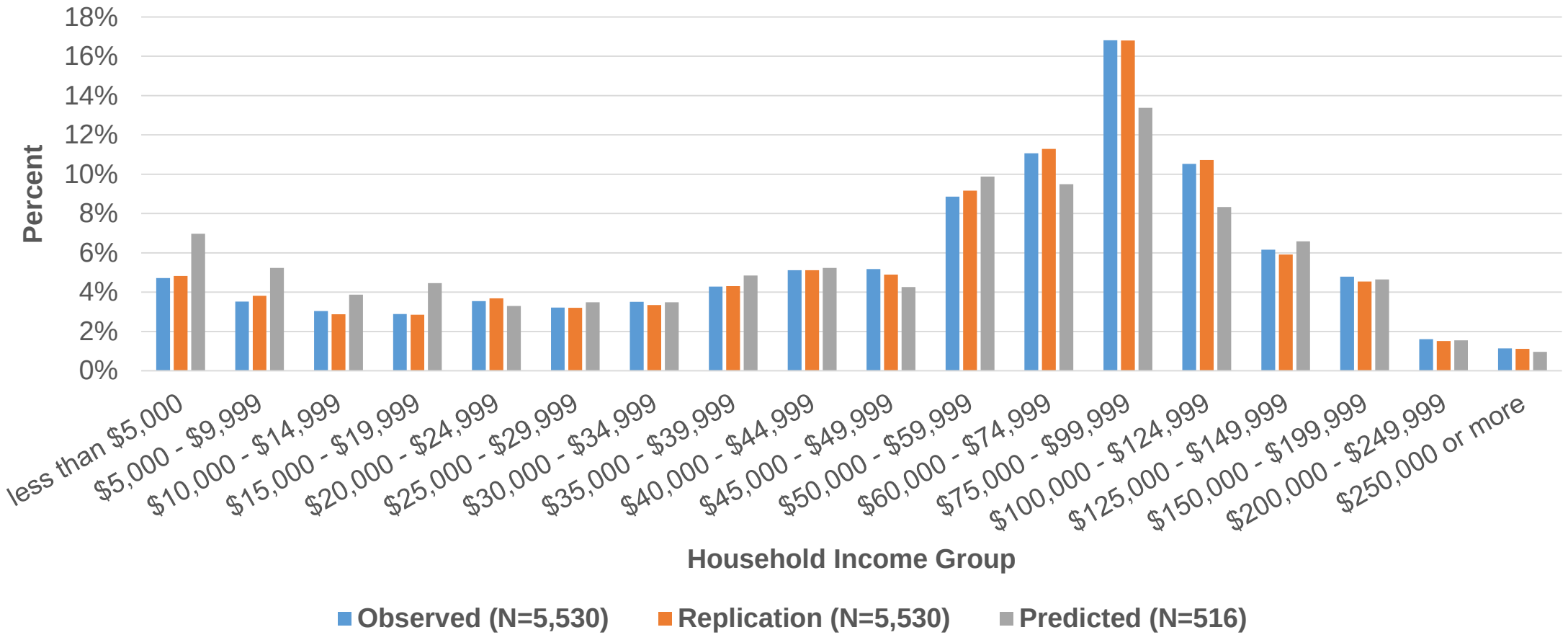
Person Variables

Geographical Level	Variable Name	Category
SuperPUMA	Employment Status	1 = Employed, 2 = Not Employed
County	Gender	1 = Male, 2 = Female
	Age	1 = 0 to 17, 2 = 18 to 24, 3 = 25 to 44, 4 = 45 to 64, 5 = 65 and Above

HH Income Imputation

- A Multinomial Logit (MNL) model was estimated to impute income for households that didn’t report income data
- Explanatory variables in the MNL model include:
  - Employment and Education Levels of Household Members
  - Lifecycle Indicators
  - Household Vehicle Ownership
  - Hispanic Indicator
  - Income Distribution of Proximal Households

Result of HH Income Imputation



- Missing data more likely in the case of low income households (see “Predicted” distribution)

Computation of HTS Expansion Weights

- Used synthetic population generator, PopGen 2.0, to estimate the HTS Expansion Weights
- PopGen 2.0 consists of two main algorithms
  - **Iterative Proportional Fitting (IPF):** Enables estimation of a joint distribution across multiple dimensions of control variables of interest
  - **Iterative Proportional Updating (IPU):** Enables estimation of weights for the sample households such that the weighted sample replicates the population frequencies in the expanded joint distributions (obtained from IPF step)

Result of HTS Expansion Weights (Maricopa)

Household

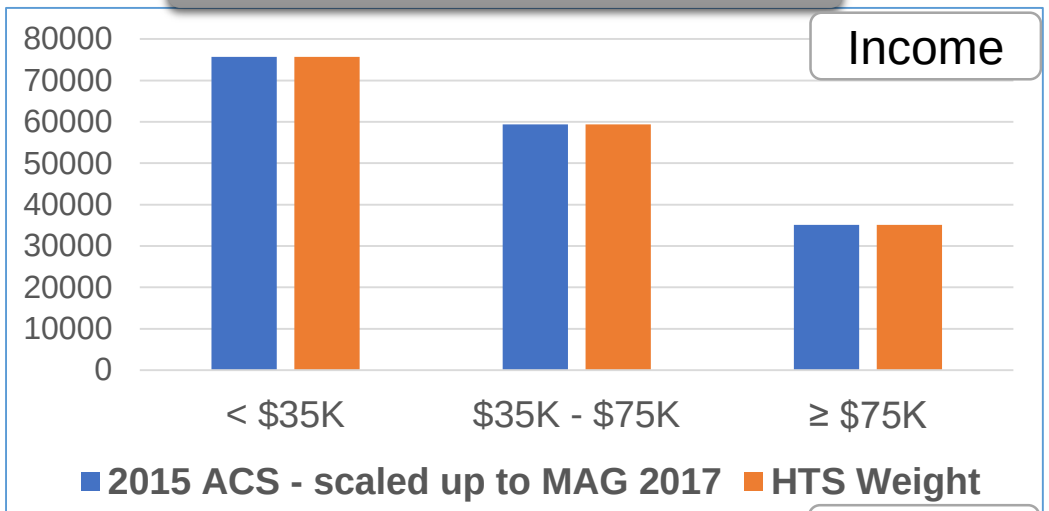
Variable Name	Variable Category	Marginal Control	HTS Weight	Difference
Household Size	1	412119	412119	0
	2	492325	492325	0
	3	247866	247866	0
	4+	406162	406162	0
Household Income	1	469110	469110	0
	2	512926	512926	0
	3	576437	576437	0
Household Race	1	252513	252538	-25
	2	1033115	1033123	-8
	3	91669	91660	9
	4	181175	181152	23
Child Presence	1	1049594	1049079	515
	2	508878	509394	-516

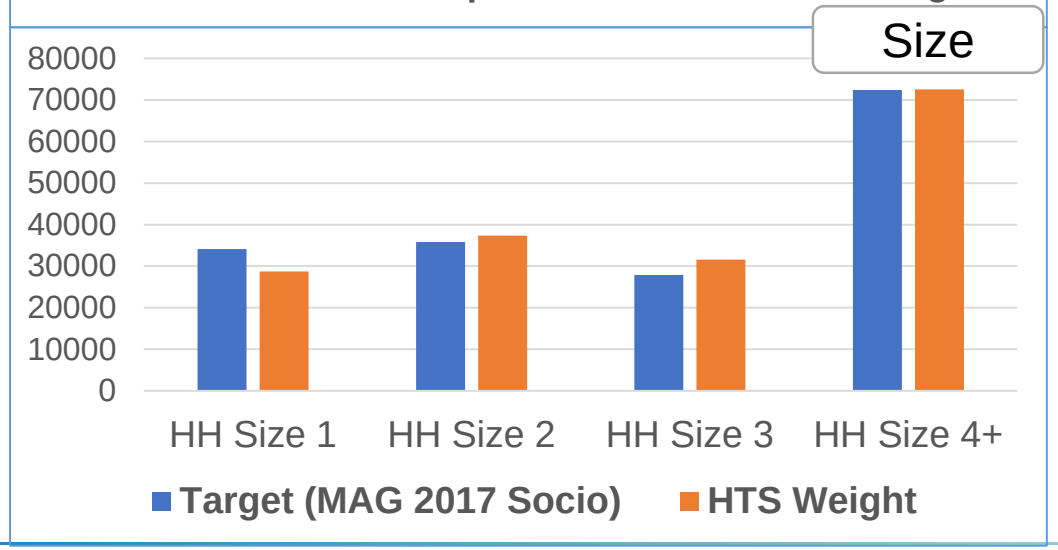
Person

Variable Name	Variable Category	Marginal Control	HTS Weight	Difference
Age	1	1171049	1171064	-15
	2	336963	336962	1
	3	1120994	1120985	9
	4	992626	992622	4
	5	543132	543132	0
Employment Status	Yes	1956066	1956067	-1
	No	2208698	2208698	0
Gender	Male	2048764	2048764	0
	Female	2116000	2116001	-1

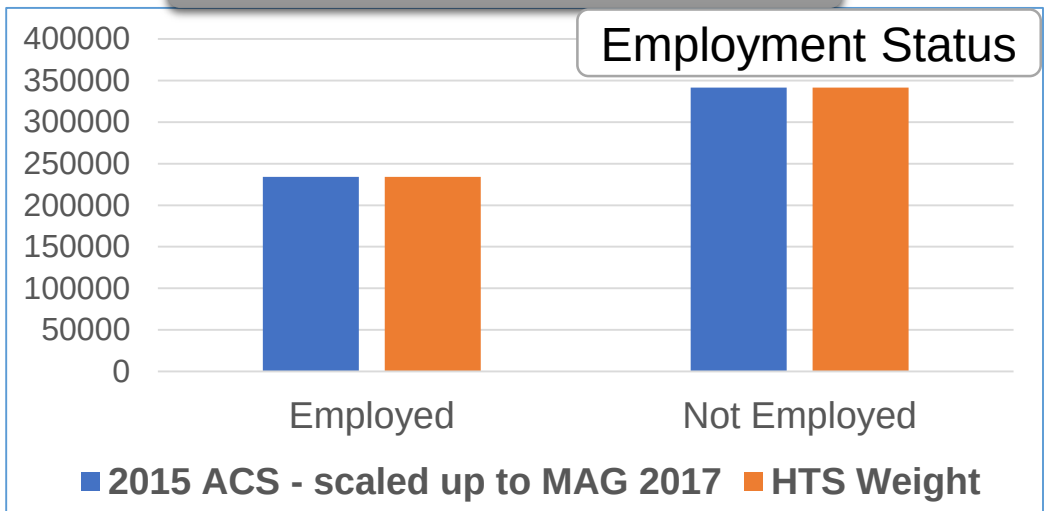
Result of HTS Expansion Weights (SuperPUMA: 121125)

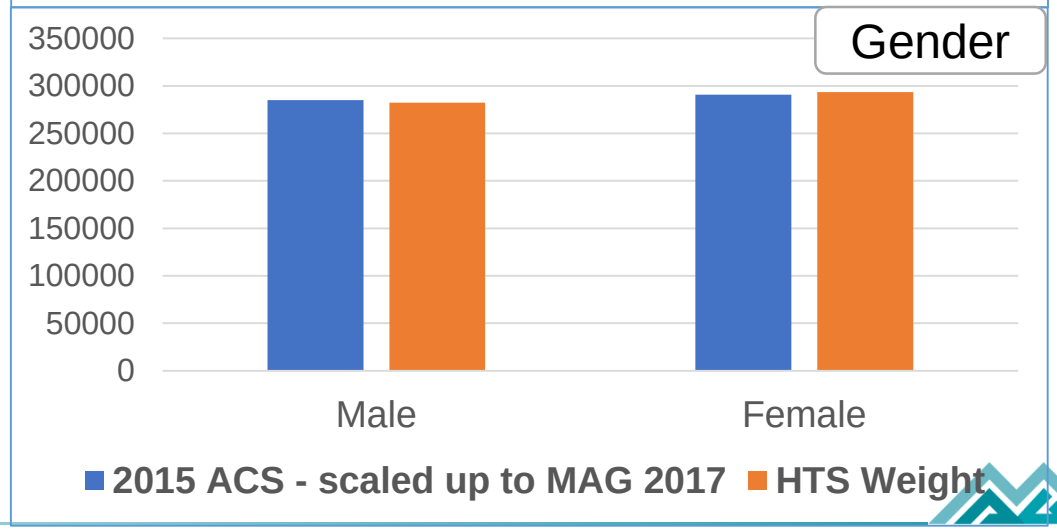
Household





Person





Conclusions

- Impute household income to minimize loss of sample prior to execution of the weighting and expansion process
- Need to delicately balance **geographic resolution**, **number of control variables and categories**, and **sample size** to avoid excess zero cells
- PopGen 2.0 based weight estimation process yielded a weighted survey sample replicating marginal control distributions for **household** *and* **person** control variables
- Using multiple geographical levels (say, PUMA & County) enhances robustness and accuracy of weighting process