SMART GROWTH AND REGIONAL COLLABORATION

Massachusetts Regional Labor Force and Household Projections
Prepared for MassDOT Office of Transportation Planning
October, 2018

Introduction

In October 2018, MassDOT, MAPC, and University of Massachusetts Donahue Institute (UMDI) completed the preparation of new population, household, and employment projections to the year 2040 to support regional transportation planning and transportation project planning and design. Population and employment projections were developed by UMDI's Population Estimate Program while household and labor force projections were developed by MAPC, under contract to MassDOT.

This document describes MAPC's methods for the development of labor force and household projections for Massachusetts regions, municipalities, and transportation analysis zones (TAZs.)

Broadly speaking, MAPC's methods and products have the following characteristics:

- Based on UMDI projections by age and gender, a model projects the number of labor force participants (workers) and households for each RPA region and forecast year;
- Using current population characteristics, the model projects household distributions by household size, number of workers, and income, at the regional level;
- Given a specified number of households in a TAZ in a future forecast year, a model provides the number of households by size, workers, and income categories at the TAZ level, which in the aggregate are consistent with the regional projections;
- Given a specified number of households in a TAZ in a future forecast year and population by age at the municipal level, the model provides the population by age categories needed for transportation planning at the TAZ level (0-4, 5-14, 15-18, 19+) and ensures consistency with the school aged population at the municipal level;
- The models and methods ensure internal consistency of the results; e.g., ensures estimates of workers in households are consistent with the labor force estimates;
- The products can support the work of other state agencies and other regional planning applications through the inclusion of projections that include other household attributes, such as household type and detailed labor force characteristics;
- The projection models can be modified through a variety of assumptions—about education levels, wages, headship rates, housing type preferences—to create a range of internally consistent scenarios for the future.

Approach

MAPC's technical approach to this work combined three major components. The first component entails estimation of future workers and households—and the characteristics thereof—from the forecasted population by applying rates and distributions derived from current Census microdata. Our multipliers are derived from the Census Bureau's Public Use

Metropolitan Area Planning Council | 60 Temple Place | Boston, Massachusetts 02111 | 617–933–0700 | 617–482–7185 fax | **mapc.org**

Microdata Sample (PUMS), which allows for more detailed cross tabulation than standard American Community Survey tables. Furthermore, these methods allow for MAPC to create and evaluate different scenarios that may arise as a result of changing economic conditions, personal preferences, or policy interventions. This was the approach taken for MAPC's 2014 population projections, which provide a wealth of insight into the relationship between labor force, household formation, and housing demand in the region.

The second component is designed to ensure internal consistency of population, household size, workers, and labor force projections. An automated process reweights the ACS Public Use Microdata Sample (PUMS) data to match a series of specified targets such as households by age of householder and household type. This method ensures consistency with projections created in different parts of the model. For example, the number of households by number of workers is adjusted to be consistent with the total number of workers projected by a more detailed labor force model.

The third component is designed to convert TAZ-level household counts into number of households by size, workers, and income that serve as inputs to the regional travel model. This component was developed by MAPC and its subcontractor Manhan Group in 2016, and adapted for this particular application. In this process, targets at a smaller geography are adjusted through a self-calibration method to ensure consistency with previously derived control totals at a larger geography. For example, the distribution of household sizes within each zone is automatically adjusted to be consistent with the regional forecasts of households by size. Furthermore, the nested structure of the model precludes the creation of infeasible household combinations, such as one person households with two workers. This method can ensure that assumptions across multiple domains—population, labor force, households—can be adjusted while ensuring that all outputs are internally consistent, an absolute necessity for travel modeling purposes.

Each of these components is described in more detail throughout the following sections.

Labor Force and Household Model

The first component of MAPC's work uses UMDI's regional population by age as inputs to a model with two distinct sub-models: a labor force sub-model and a household sub-model. Both rely on the Census Bureau's Public Use Microdata Sample (PUMS) data to create multipliers applied to future projected populations. Please read the sections below about the data inputs to this model and reference the diagram at end of this document for more details about these sub-models.

PUMS Geographies
PUMS data-essentially individual survey responses-are associated not with specific tracts or block groups, but with geographies called PUMAs that must contain at least 100,000 people, in order to protect privacy. PUMAs almost always contain multiple municipalities; they can change over time, and they do not adhere to RPA boundaries. Please refer to Appendix A for the lookup tables of 2000 and 2010 PUMAs to RPAs.

Household and Group Quarters Population by Age
The CTPS model requires the following population inputs: total household population, household population age 0-4, 5-14, 15-18, and 19+, population in dormitories, and

population in other group quarters living situations. Because of this, MAPC's household model requires projections—at both the regional and municipal level—of population in households by 5-year age groups age, as well as more detailed estimates of the population age 15-17, 18, and 19 years old in order to break the population into age categories for CTPS.

Decennial census data do not provide living arrangement by single year of age; the tables that include both living arrangements and age combine the 18 and 19 year old population into a single five-year category. Therefore, a process is required to combine information about living arrangements by aggregate age groups with single-year age counts. The following process was used for each municipality and aggregated to create a total population by age and living arrangement for each region.

1. Using Census 2010 SF2 table PCT3 (Sex by age [single year], Total population), disaggregate the 15-19 year old total population (year 2010), by age and sex, into population 15-17, and population 18-19.
2. Also using table PCT3, calculate the percent of 18 &19 year olds who are age 19.
3. Combine PCT3 with SF2 table PCT 5 (Sex by Age [five year groupings] for the Population in Households) to determine the percent of all age groups (including the 15-17 and 18-19 subgroups), by sex, who are in households vs. group quarters
4. Use the results of step 2, disaggregate the 18-19 year old population in households into single year age groups. (this assumes that the 18 : 19 year old age ratio of population in households is the same as the overall population)
5. Combine the results of the above steps to calculate the share of the 15-19 year of population in households, falling into the following three groups: age 15-17, age 18, age 19. These proportions were applied to the household population by 5-year age groups for future years.

For all other age categories the group quarters population was broken out using Census 2010 SF2 table PCT5. We applied the 2010 share of population in group quarters, by five-year age group, to the future year projected population by age group.

Household Civilian Population by Age

MAPC's Labor Force model excludes those and group quarters and those in the armed forces who live in households. The count of armed forces workers in households is not a large number in Massachusetts and not expected to change dramatically in the future. Therefore for each region the total armed forces population living in households by age and sex from the 2008-2012 PUMS was carried forward in future years. This number was subtracted from the total household population by age and sex to create a civilian household labor force by age and sex.

Educational Attainment and Labor Force

Labor force projections were produced by applying age-specific labor force participation rates to the projected population in households. Following consultation with MassDOT and the Massachusetts RPAs, MAPC used 2012-2016 PUMS data to develop rates for the labor force model. Labor force participation rates for older adults have increased substantially since 2000; the use of the most recent rates captures these trends, but does not assume any further increases (or potential decreases) in labor force participation among older adults in coming decades. This could have the effect of underestimating future labor force if post-65 employment becomes relatively more common than it is today.

The likelihood of each age category achieving a given educational attainment level was applied to the future population by age categories. Age categories for determining rates were 15-19, 20-24, 25-34, 35-44, 45-54, 55-64, 65-74, and 75+. We did not determine rates at the 5-year age category level due to high margins of error, especially in smaller RPAs. Educational attainment levels were classified as "High School or less", "Associate's degree or some college", "Bachelor's Degree", and "Master's or higher".

To estimate future labor force, MAPC applied these age-specific educational attainment rates to the projected civilian population in households to yield population by age, sex, and educational attainment. Labor force participation rates (again, based on 2012 -2016 PUMS data) specific to age and sex categories were applied to yield the number of residents in the labor force.

Headship Rates
The models specify three different types of households: single person household, 2 or more person household with children under 18, and 2 or more person household without children under 18. Following consultation with MassDOT and the Massachusetts RPAs, MAPC used 2008-2012 PUMS data to develop headship rates by household type. Headship rates of younger adults declined between 2008-2012 and 2012-2016, most likely because of the extremely tight housing market across the state. The group decided to plan for the household demand associated with the 2008-2012 housing market.

The likelihood of each age category being the head of a household or a member of each of the household types was determined for each of the following age categories: 15-19, 20-24, 25-34, 35-44, 45-54, 55-64, 65-74, and 75+. We did not determine rates at the 5-year age category level due to high margins of error, especially in smaller RPAs. These rates were applied to the projected civilian population in households in order to produce households by type and age of householder, as well as population by age in each of the three household types.

Outputs
After applying the rates described above to UMDI's future population by age and sex, this method produces regional population by age and household type, household by age of householder and household type, and labor force by age, sex, and education level.

Household Detail Model

For the regional household projections to serve as useful control totals for the TAZ projections that are input to the travel demand model, more work is needed to create additional detail about households, namely distribution of households by size, workers, and income. All of this detail must be consistent with the previously produced regional totals.

In order to achieve the necessary level of internal consistency, MAPC used a new analytical tool currently referred to as a PUMS 'reweighter'. The 'reweighter' is a method created by Alan Clayton-Matthews, Associate Professor of Economics and Public Policy at

Northeastern University and Senior Research Associate at the Dukakis Center. In collaboration with Dr. Clayton-Mathews, MAPC adapted the technique to this particular application. This technique adjusts the household weights of PUMS data so that the resulting household population matches a set of user-defined targets. The user can specify known characteristics of future residents/households and use the method to determine what more detailed characteristics of a synthetic future population would be. This method ensures consistency with projections created in different parts of the model. For example, the number of households by number of workers is adjusted to be consistent with the total number of workers projected by the labor force model described above.

The 'reweighter' has three input tables which act as targets for the future reweighted sample to hit: households by householder age and household type, population by age and household type, and labor force participation by education level and age. Age categories for the input tables are classified as 5-year age cohorts. The household types match the three household types created in household model, and the education levels correspond to the four education levels developed in the labor force model.

The microdata sample to be reweighted are the 2008-2012 PUMS data separated by the PUMAs that make up each RPA (see Appendix A). However, unlike the PUMA geographies used in the labor force and household models, the reweighter only used one PUMA which covers Nantucket for NPEDC which allowed for a better model fit.

The method is run on each RPA individually. The records are reweighted by an application written in the programming language R. The procedure operates on person level records in a particular household. The weight of all members of a household is adjusted by +1% or -1%. If the change (either positive or negative) leads to the overall sample becoming more similar to the target input tables, the change is "accepted" and applied to all members of that household; if the change results in greater divergence of the synthetic population and the target values, the change is rejected. The output of this method is an entirely reweighted PUMS sample of the RPA which when aggregated and weighted, approximates the targets developed by the regional household and labor force models. These results ensure internal consistency with new regional statistics and allows us to investigate detailed characteristics of the future population.

This method produces a dataset that is remarkably close to every one of the input targets. Even in FRCOG, an RPA with only one PUMA of records, the reweighted PUMS dataset had an average variance of only 0.04% from each of the specified targets of population by age and household type. Similarly, the reweighted PUMS sample had an average variance of only 0.2% from each of the specified targets of households by householder age and household type. The labor force targets were not met as closely, with an average variance of 2.9% for each of the targets of labor force members by age and education level. It is not surprising that these targets are more difficult for the 'reweighter' to match because the targets for labor force were created using multipliers from 2012-2016 PUMS data as opposed to 2008-2012 PUMS data. The largest percent variances were found in small categories— population over 75 in the labor force at certain education levels. Excluding any labor force targets for the population over 75, the average variance is 0.0%.

This reweighted PUMS sample for each RPA was used to develop regional targets of households by income level, number of workers, and household size which are used for transportation modeling purposes.

Third Technique: TAZ level 'Disaggregator'

To produce the TAZ-level household details needed for transportation planning purposes, MAPC created a method to disaggregate total households by size, workers, and income, while ensuring that the sum of TAZ projections matches the regional detail produced by the methods described above.

Using municipal- and age-specific headship rates, MAPC developed total household estimates for each municipality and adjusted the total to match the RPA total households described above. The resulting household totals for each municipality were allocated to constituent TAZs by the relevant RPA. Further detail on households in each TAZ was created through Iterative Proportional Fitting (IPF) designed to estimate households by householder age and household type in each TAZ, consistent with regional control totals. The 'seed' of this IPF was created for each future year and each RPA by the reweighted PUMS sample explained above. Total households were assigned an initial joint distribution to age and household type based on the reweighted PUMS sample for the PUMA containing the TAZ, and then the distributions were adjusted for all TAZs in an RPA so that the regional control totals by householder age and type (jointly distributed) were satisfied. MAPC then cross-classified these households by size, workers, and income categories using a custom-built 'disaggregator,' (described below) which produced the marginal values that serve as inputs to the CTPS travel demand model.

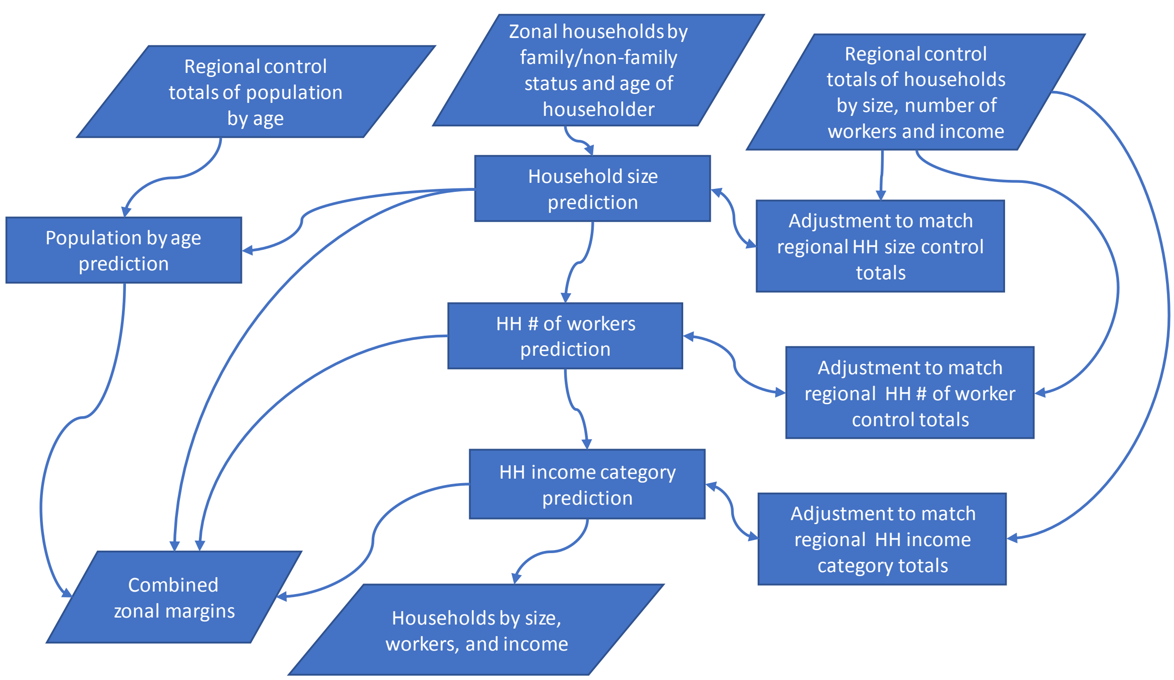
Disaggregator Framework
The custom-built disaggregator used for this work is designed to classify households (by householder age and type) into even more detailed categories of size, workers, and income, in a manner that is both logically feasible and consistent with the regional control totals for each category of households. The disaggregator consists of a series of multinomial logit models that predict household membership in a specific category for each one of the three pertinent dimensions of household type: size (1, 2, 3, 4, or 5 persons), number of workers (0, 1, 2, 3 or more), and annual household income (<$35K, $35-75K, $75-125K, $125K+). A final post-process predicts population by age from households by size using static shares derived from PUMS data.

There are natural relationships between these levels: a household's size constrains its maximum number of workers, and intuitively a household with fewer workers may be expected to earn less income (or, at the very least, a household with no workers can earn no income in the form of salary wages). Therefore information from the result of running the size category membership model can be passed to the number of workers model (and the population by age model), as well as from the number of workers model to the annual household income model, to ensure internal consistency of results.

At a more regional level, control totals of households by size, number of workers, and income have been established for each RPA through the PUMS 'reweighter'. The disaggregator models include a so-called "auto-calibration" routine that summarizes the sub-model results to compare with these regional control totals and iteratively adjusts the alternative-specific constants of the models in order to achieve as close a match as possible. This adjustment process happens at each level of the hierarchy before continuing to the next. While the disaggregator produces fully cross-classified households (age x type x size, x workers x income), consistency with the regional control totals for size, workers, and income is achieved for each dimension, not for a joint distribution. The results of the disaggregator have an average variance of 0.00% from the targets by size for each RPA. Similarly, the average variance from the targets by workers and targets by income were 0.05% and 0.02% respectively, on average for all RPAs.

This fully cross-classified table of households by size, workers, and income could be used in place of the output currently generated internally within the CTPS travel model by an iterative proportional fitting (IPF) algorithm, as part of the trip generation process. However, in order to remain compatible with current business processes, the disaggregator also provides marginal totals by household size, workers, and income, that can be used as input to the IPF process. These are required to be integer values; therefore a "bucket-rounding" process was used to convert the floating-point numbers produced by the disaggregation process to whole number totals while remaining as faithful as possible to the probabilistic distribution of the underlying data.

To create the integers needed, the cross-classified table is summed across categories to create zonal totals for each size, worker, and income category. The four categories across one attribute (i.e. size) are added together to determine the total number of households in a zone. This number is rounded to the nearest integer. The floating-point number for each of the four categories in this attribute (i.e. 1 person, 2 person, 3 person, 4p person) is rounded down to the integer below its value. The integer categorical totals are added together for each zone and compared to the integer total number of households in each zone. The remainder households (between 0 and 4 households) are assigned to one of the four categories based on category's deviation from the original floating point value in each zone after rounding.



Estimation
Census 2007-2011 ACS PUMS data were used to estimate the multinomial logit models for predicting household size, number of workers, and income category. The "multinom" function included in R's *nnet* library was employed to estimate the model. The model contains only minimal household characteristics and location-specific dummy variables by PUMA.

Certain dimensions of household type do affect availability of alternatives; for example, the household type scheme separates singles from couples and families, and household size affects number of workers. The decision was therefore made to reflect alternative availability explicitly in the code implementing the disaggregator, as discussed in the section below on implementation.

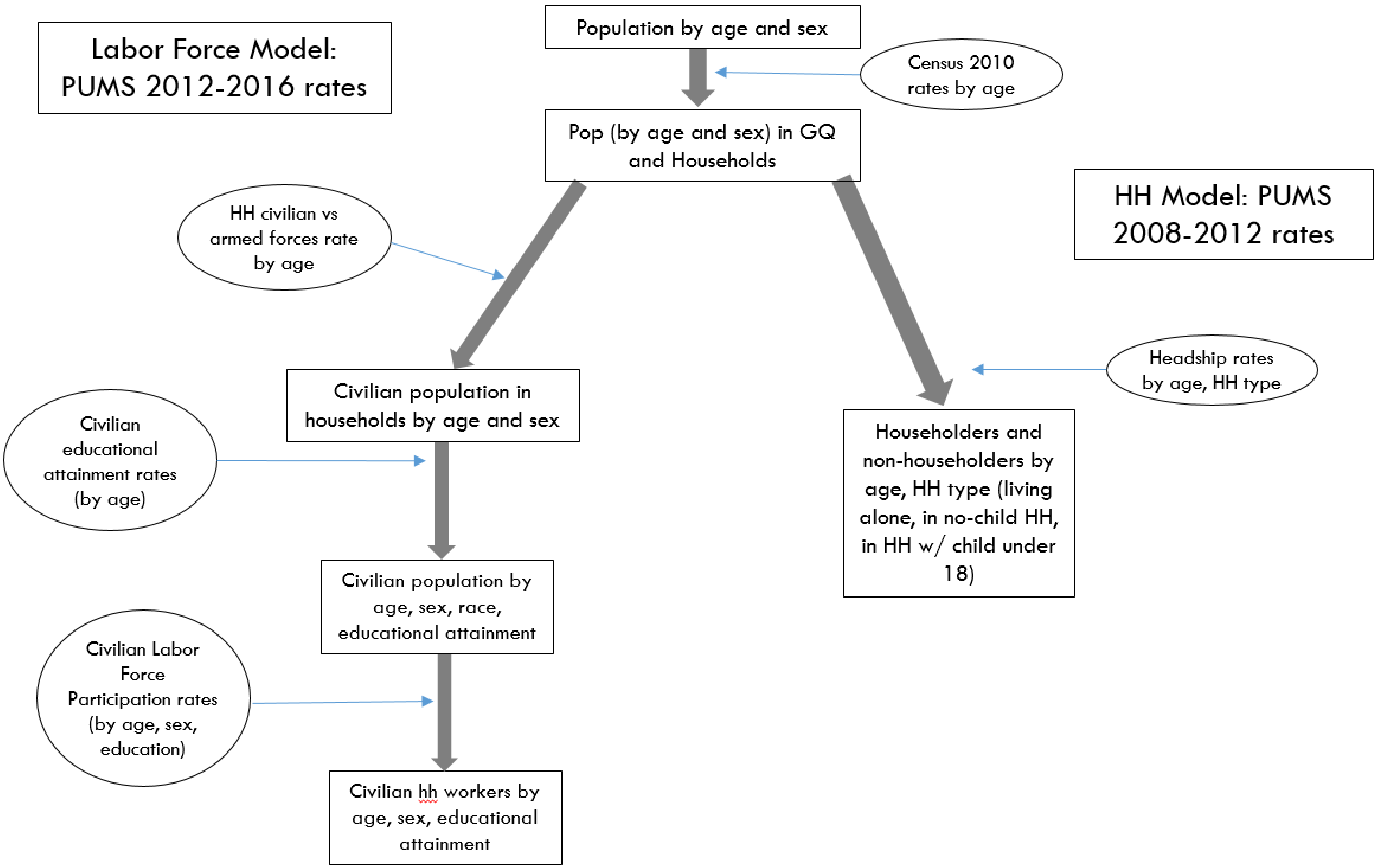
Population Adjustments
In addition to producing detailed household estimates at the TAZ level, the disaggregator also estimates population by age for each TAZ, and seeks consistency with the pre-existing municipal population projections produced by UMDI. Because the population by age sub-model works differently from the three main levels of the household type hierarchy, the way in which it achieves correspondence with regional/municipal control totals works differently as well. Fixed shares (derived by MAPC staff from PUMS data) are used to obtain initial estimates of population by age category based upon the distribution of households by size in each zone. This initial distribution is then adjusted to match input municipal-level control totals of population by age group. Total population in each zone is then checked for logical conditions that must be met such as:

- $pop_i \geq hh_{size1,i} + 2 * hh_{size2,i} + 3 * hh_{size3,i} + 4 * hh_{size4,i}$
  (minimum population requirement)
- $pop_i \leq hh_{size1,i} + 2 * hh_{size2,i} + 3 * hh_{size3,i} + 12 * hh_{size4,i}$
  (maximum population constraint, based upon highest observed size of households with 4+ persons)

Implicitly, these checks also ensure no zones in which there is population but no households nor vice versa. Zones with no households having four or more people also cannot be adjusted in population.

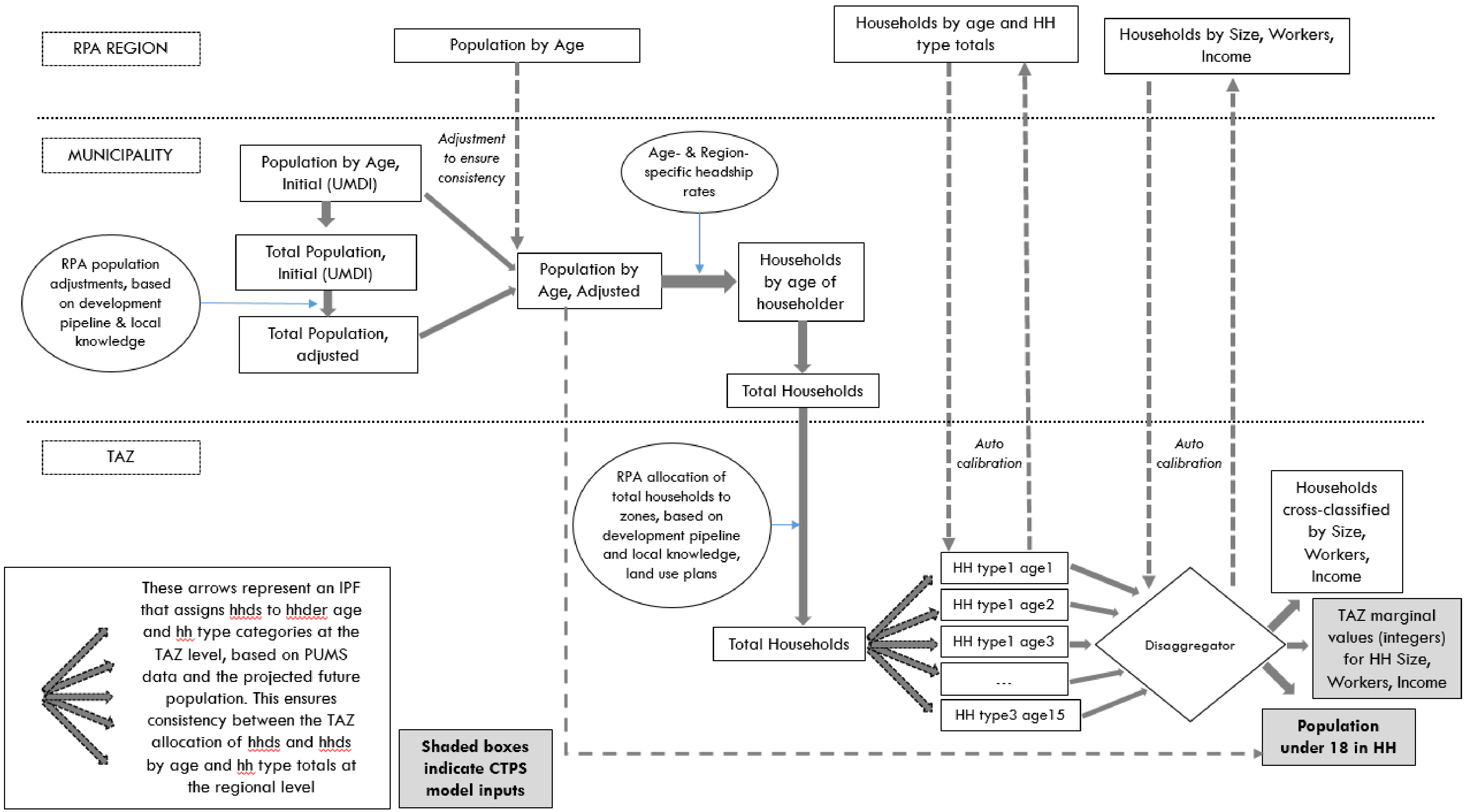
Because household size is determined by zone before the total population by age, the model was unable to match the total municipal population targets in every municipality. In these cases, MAPC staff made sure the child-age population (0-4, 5-14, 15-18) matched the municipal control totals and the excess population was moved to the 19+ category. This was done because the CTPS travel demand model uses the child aged population to model the number of school trips while the total 19+ population in each zone is not used for trip generation purposes.

Massachusetts Regional Household Projections: Model Schematic
September, 2018



MAPC

Massachusetts Municipal and TAZ household projections
Model Schematic
September, 2018



Appendix A: PUMA RPA Lookups

MAPC 2010	MAPC 2000	OCPC 2010	OCPC 2000	SRPEDD 2010	SRPEDD 2000	NMCOG 2010	NMCOG 2000	MVPC 2010	MVPC 2000	CMRPC 2010	CMRPC 2000	FRCOG 2010	FRCOG 2000
503	400	4903	4600	4902	4100	501	500	703	900	300	1500	200	200
504	900	4000	4000	4200	4200	502	600	701	700	302	2100		
505	1000	4901	4100	4301	4300	1300		702	800	303	2200		
506	1100	4902		4302	4400					304	2300		
507	1200			4303	4500								
508	1300			4500									
703	1400												
704	2400												
1000	2500												
1400	2600												
2400	2700												
2800	2800												
3301	2900												
3302	3000												
3303	3100												
3304	3200												
3305	3301												
3306	3302												
3400	3303												
3500	3304												
3601	3305												
3602	3400												
3603	3500												
3900	3600												
4903	3700												
	3800												
	3900												

PVPC 2010	PVPC 2000	BRPC 2010	BRPC 2000	CCC, MVC, NPEDC 2010	CCC, MVC, NPEDC 2000	MRP C2010	MRPC 2000
1600	1600	100	200	4800	4700	301	400
1900	1700		100	4700	4800	400	300
1901	1800						
1902	1900						