Massachusetts Regional Housing Demand Projections
Technical Methodology
Prepared by the Metropolitan Area Planning Council
January, 2025

Introduction

In Winter 2025, the Massachusetts Executive Office of Housing and Livable Communities (EOHLC), the University of Massachusetts Donohue Institute (UMDI), and the Metropolitan Area Planning Council (MAPC) completed new population and household projections out to the year 2050 to support EOHLC's State Housing Plan. Under contract to EOHLC, UMDI created population projections at the regional level. MAPC then developed household and housing demand projections based on these population projections. The resulting products are four housing demand scenarios that helped EOHLC develop a housing production goal for 2035.

This document details MAPC's methods for the development of regional household projection scenarios. UMDI's methods for the creation of regional population forecasts are described elsewhere in documentation prepared by UMDI.

A simplified version of MAPC's projections process is as follows:

- Based on UMDI's regional population projections, MAPC first determined the number and percent of people in households by age group for each Regional Planning Agency (RPA) geography and forecast year using values computed using data from the U.S. Census Bureau's 2020 Decennial Census. These estimates made assumptions about the likelihood that someone of a given age group was in group quarters. For the 15-19 and 20-24 year old age groups, MAPC assumed the number of individuals in group quarters remained constant across all projection years. For every other age group, MAPC assumed the percent of people in group quarters remained the same for all projection years.
- For each scenario, headship rates (the likelihood of a person of a particular age category being the head of a household) are calculated. More detail about how these rates were created for each scenario are below. In addition, assumptions about the housing needed for the Massachusetts family shelter population were added in for each scenario.
- Following completion of household estimates, household population counts were reviewed and adjusted to ensure consistency with the regional UMDI Population projections for each age group. This process entailed adding or removing people from selected microstimulated households to meet population targets without changing essential household characteristics.
- Housing unit demand was estimated for each scenario based on projected households, scenario vacancy rate assumptions, and an adjustment for the loss of built units to seasonal uses.

Household Population

Following receipt of the regional population projections from UMDI, MAPC's first step was to estimate the population in households for both 2020 and all forecast years. Population in households excludes military personnel and residents living in group quarters (dorms, correctional facilities, nursing homes, etc.) MAPC adhered to the definition of "household" used by the Census Bureau (more information is available in Census Bureau technical documentation) which is included as a variable in the PUMS data (see the PUMS data dictionary). To convert the UMDI data, MAPC first determined the proportion of the population living in households as a share of total population by age group and RPA. Using data from the 2020 Decennial Census DHC file, MAPC then divided the number of people in households (which excludes people who live in group quarters, including military personnel and students) by the total population by age and RPA. For the 15-19 and 20-24 year old age groups, MAPC assumed the number of individuals in group quarters (mostly dormitories for this age group) remains constant across all projection years, since. For each other age group, MAPC assumed the percent of people in households remains the same for all projection years. The resulting rate is the proportion of the population in households by age and RPA in 2020. To produce the projected household population, the household population proportions were applied by age and MPO to produce the projected population in households for each projection year. For the 15-19 and 20-24 year old age groups a constant number of group quarters was subtracted from the projected population in each projection year.

Regional Households

Once the population in households had been determined, the next step was to estimate the number of each type of household likely to form in each region for each scenario. Households were classified by the age of the householder (i.e., the head of household as defined in Census data), number of people in the household, and presence of children under age 18. MAPC's work achieved these ends using the U. S. Census Bureau's American Community Survey (ACS) Public Use Microdata Sample (PUMS) data to construct rates which were applied to the projected population in households. The following sections provide a detailed account of how these rates were created, implemented, and adjusted to create the final housing demand scenarios.

ACS PUMS Geographies

Household rates for each region were based on PUMS data. Each PUMS release covers a five-year period and is available at a geographic level called a Public Use Microdata Area (PUMA) which contains at least 100,000 residents. This is done to maintain the privacy of ACS respondents, since the data made public in PUMS contains specific information about individual persons and households. PUMAs are usually larger than municipalities and generally do not adhere to county or regional boundaries (with a few exceptions). The geographies of PUMAs also change every decade; PUMAs boundaries for 2000, 2010, and 2020 are not always aligned with each other. Therefore, MAPC constructed crosswalks that assign each decade's PUMAs to MPO regions. PUMAs for each region were then pooled to create a regional group of PUMS records used for household and labor force estimation. Please refer to Appendix A for the PUMA 2000 and 2010 to MPO crosswalks.

Household Types

For these projections, MAPC defined three general Household Types: a person living alone, households with two or more adults and no children, and all households with one or more children under 18. These types were chosen because they are assumed to have very distinct housing needs and household location choice preferences. Following consultation with EOHLC and UMDI, MAPC developed four scenarios with different region- and age-specific headship rates (described below), defined as the probability of a person of a given age being the head of a given type of household. Headship rates for the three groups of households described above were applied to the total population in households for each forecast year. The resulting product is an initial estimate of households by age of head of householder and household type for each region and forecast year (2025, 2030, 2035, 2040, 2050). These household forecasts were used as targets in the PUMS "reweighter".

PUMS Reweighting

After determining the number of households, additional work was needed to create detailed household estimates that allow EOHLC to understand household characteristics by size, number of workers, and income. This detail was created using a process called "PUMS reweighting." 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. This technique adjusts the household weights of PUMS data so that the resulting collection of weighted households matches a set of user-defined targets. The user can specify 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 a more detailed explanation of how the PUMS reweighter works see the UMDI employment projections documentation. MAPC used two input tables for the 'reweighter' which acted as targets for the future reweighted sample to hit: households by householder age and household type, and households by householder age and tenure. Age categories for the input tables were classified as 5-year age groups. The household types matched the three household types created in the household model. The microdata sample to be reweighted were the 2017-2021 ACS PUMS data grouped by the PUMAs that make up each RPA (see Appendix A).

The outputs of the "reweighter" were a list of PUMS "serial numbers" (unique identifiers for households), and the adjusted household weights for each forecast year. Those were merged to the original 2017-2021 ACS PUMS records. MAPC then created a cross tabulation of the total number of households by the number of people in the household, presence of children in the household, the age of the head of household, tenure, and the household income group of the household for each year and RPA region.

Scenarios

As mentioned earlier, four scenarios were created for this planning process, all of which have different assumptions about headship rates, vacancy rates, and the treatment of the current Massachusetts family population in shelters. The scenarios are distinct in their assumptions, explained below:

Headship Rates

I. **Scenario 1**: Baseline headship rates were calculated using 2017-2021 PUMS data by household type and age of householder.

II. **Scenarios 2-4**: MAPC adjusted 2021 baseline headship rates to account for latent demand in two ways including accounting for a portion of subfamilies living in households with children, and assuming a return to 2000 headship rates for households without children and heads of household under the age of 45.

a. *Households with children:* In Massachusetts there are 63,000 households with subfamilies—family units currently living in a household with another family. For a subset of households with children and subfamilies (defined below) we formed a new household for each subfamily. New households were classified based on the characteristics of the people in each subfamily, so subfamilies with children formed new households with children and subfamilies without children formed new households without children. Once these new households were formed in the data, MAPC recalculated new headship rates for households with and without children by household type. The characteristics of households for which we formed new households from subfamilies are defined below. These 15,000 subfamilies comprise a quarter of all subfamilies living in households with children in Massachusetts.

i. Subfamilies living in overcrowded non-multigenerational[1] households with children (9,000 new households)

ii. Subfamilies living in overcrowded[2] multigenerational households with children (6,000 new households)

We phased the formation of these new households over time and assumed all 15,000 subfamilies form their own household by 2035. We assumed half of those households will form in 2030, and the remaining half will form new households by 2035.

b. *Households without children:* For households without children where the householder is under 45 years old, we used 2000 headship rates as a baseline; in the case where the adjusted 2021 headship rates for households without children, accounting for newly formed households without children from the subfamily analysis above, are higher than the 2000 headship rates, we used the higher rate. Households with a householder 45 years or older retained the adjusted 2021 headship rates, whether or not the 2000 rate for their age group was higher.

[1] Multigenerational households are defined by the Census Bureau as family households with three or more generations.

[2] Overcrowded households are defined as those with more than two persons per bedroom in the housing unit.

Vacancy Rate Assumptions

MAPC calculated current vacancy rates by tenure in each region using Census 2020 data. Vacancy rates are calculated as the number of vacant units divided by the number of vacant plus occupied units. Homes that are set aside for seasonal use and those vacant for other reasons are not incorporated into the vacancy rate calculations. The projected mix of households by tenure for each region is used as the basis for calculating the number of for-sale and for-rent units needed in a forecast year. The following assumptions were used for the scenarios:

I. **Scenario 1**: Vacancy rates were held constant to the current rates by RPA and tenure. As new households were added in future years, we added enough vacant units to maintain the current rate.
II. **Scenarios 2-4**: Rental and owner target vacancy rates are set at 7.4% and 1.5%, respectively, with the aim of achieving a combined average vacancy rate of roughly 4.5% across all units. The actual combined vacancy rate will vary based on the proportion of rental and owner units in each RPA.

Family Shelter Population

For scenarios 2 and 3, MAPC added onto projected households starting in 2035 the number of households currently living in Massachusetts' Family Shelter System using data provided by EOHLC. We received two datasets from EOHLC: 1) the number of Emergency Rental Assistance applications by RPA and 2) the number of people and the number of heads of household in the Family Shelter system by 5-year age group. To translate the number of people and households from EOHLC into the appropriate inputs to the projection's workflow, MAPC estimated the number of shelter households in each RPA and created a synthetic household record for each shelter household using the ACS PUMS.

First, MAPC distributed the shelter households to each RPA. To do so, we used information about the municipal distribution of current Emergency Assistance (EA) applications which account for every family applying for emergency family shelter. We assigned an RPA designation to each municipality and summed the EA applications to the RPA level. From there we calculated the proportion of EA applications in each RPA and applied that percentage to distribute shelter population in households to RPAs. Since the objective of this analysis is to determine how many housing units would be need to accommodate the current shelter population, we assume all residents are in households and none are in group quarters.

Then we determined the number of shelter households in each RPA. We calculated the headship rate for each 5-year age group by dividing the number of shelter householders by the total shelter population in that age group. We then multiplied the shelter population in households (by 5-year age group and RPA) by the shelter headship rate to find the number of households in each 5-year age group in each RPA. Next, we used the number of households by RPA and 5-year age group as targets to sample PUMS records to create a synthetic household record with person and household characteristics for each household in the family shelter system.

To narrow the full PUMS to a sample which reflects the shelter population, we made assumptions about household characteristics, including that all shelter households will be renters and earn less than or equal to $35,000 annually. To create a subset of PUMS records which approximate the characteristics of shelter households, we randomly sampled individual PUMS householder records for renter households with a household income under $35,000. We exactly matched the number of households in each 5-year age category and RPA. A single sample, however, may not match the total number of people in shelter households by age. To remedy this, we sampled PUMS data from each RPA 500 times and chose the sample which most closely matched the population in households by age from the shelter system data.

Once we sampled the household records, we used the serial number of the PUMS household to append the person records to the sample. The final sample of PUMS records constituted the "shelter households" which were then appended to the post-processed reweighted PUMS records in 2035 and each subsequent five-year period.

Post-processor

As described above, the MAPC "reweighter" produces a detailed set of projections by household characteristics. However, because we did not utilize a reweighter target for the *size* of the projected households, the resulting outputs from the reweighter inevitably deviated from the projected population totals by age. The reweighter only adjusts the household weights in the PUMS records, not the person weights, so it cannot directly produce projected households whose size and age characteristics are completely in line with the population projections from UMDI. However, because PUMS records include a record for each household, with household weights, and a record for each person in each household, with separate person weights, we can use a post-processor after running the reweighter to adjust the number of people assigned to each individual household in order to bring the final household population by five-year age group in line with the UMDI projection targets. To do this, we took a sequential approach: 1. MAPC ran the reweighter to obtain projected households that match our future year targets of households by type, tenure, and age of householder; and 2. MAPC ran a post-processor that adjusted the size of these future households to align the total population in households with the UMDI projections of population by age.

The population post-processor first assessed the deviation between the UMDI population projections and the population by five-year age group in the projected households created by the household reweighter. Next, for each five-year age group, it identified a subset of households that could be adjusted to close the gaps. The goal was to adjust the number of people in certain households without changing the total number of households by household type, and without creating unusually large or otherwise unrealistic households. Since single adult households are ineligible for adjustment as that would change their household type, the post-processor focused on multiple adult households without children, and households with children.

For age groups where the projected household population was lower than the UMDI population, the post-processor added additional person-records of the matching age group to households. The additional person-records were obtained by random sampling from within the total household

population. These person-records were duplicated and added to eligible households (i.e. we do not shift person-records between households, we duplicated a person-record and assign the duplicate to another household).

For all of the adult age groups (18+), we only added person-records to households with multiple adults and no children. Households with children did not receive extra adults. For the under-18 age groups, we only added person-records to households with children where the number of adults is greater than one, and the number of children already in the households was less than four. For adult age groups we only added person-records to households with less than eight adults, and where the age of the householder is also within +/- 10 years of the age of the person-record being added in. For all cases, no household receives more than one additional person-record.

For age groups where the projected household population is higher than the UMDI population, the post-processor identified households where a person-record can be removed from the household without changing the household's type. Person records were only removed from households that were either: 1. Multiple adult households without children that have between three and eight adults, or 2. Households with children that have three or more adults (for adjusting adult population) or more than one child (for adjusting child populations). No household could have more than one person-record removed across all the age groups, and the householder person-record (SPORDER==1) was never removed.

By progressing through each age group deviation, in each RPA, in each forecast year, the post-processor adjusted household sizes up or down as described above until the deviation between the UMDI population projection and the MAPC projected household population was zero, or in a few edge cases, made as small as possible given the restrictions on which households can be adjusted as described above. At the conclusion of the post-processor, there were a small number of age group / RPA / year combinations (less than 10) where a complete convergence with UMDI population estimates was not possible. However, in all cases the post-processor reduced the deviation to no more than twenty persons across the entire RPA.

Housing Unit Demand

To estimate the number of homes lost to seasonal conversion between 2010 - 2019, EOHLC estimated how many households in 2019 were living in homes built since 2010, and then subtracted that from the net change in year-round housing units over that same period. For example, Cape Cod added 8,800 units but saw the number of year-round homes decline by 3,400 even as there were 2,400 households living in newly built year round homes. These figures indicate that 5,800 homes were converted from year-round to seasonal use over that period. Similar patterns are observed, to a lesser extent, in three other regions: Berkshire County, Pioneer Valley, and Franklin County. In all four of these regions, the number of homes converted to seasonal use was converted to a percentage and then applied to the 2020 housing stock to estimate seasonal conversion over the coming decade.