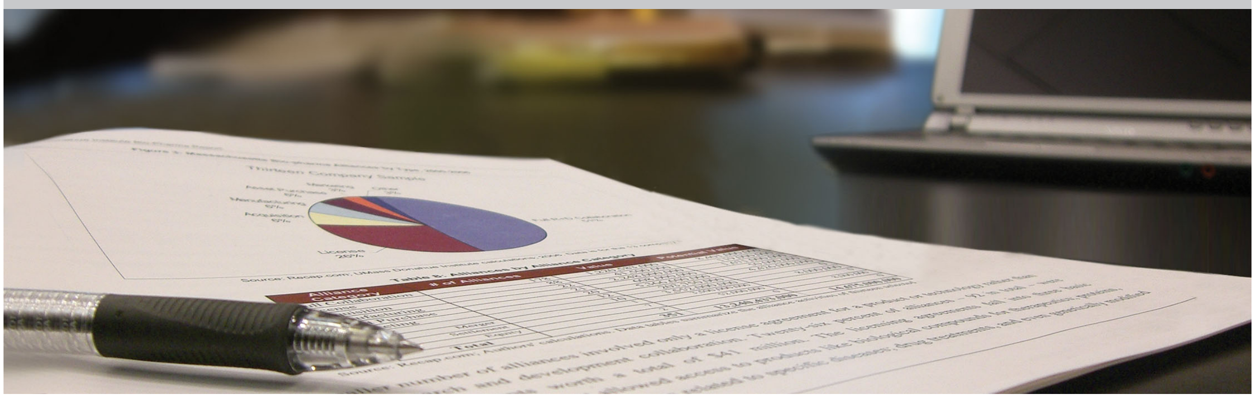
Massachusetts State and Regional Employment Projections

Overview of Methodology and Projections

June 2023



UMassAmherst | Donahue Institute
Economic and Public Policy Research

Massachusetts State and Regional Employment Projections

Prepared by the UMass Donahue Institute's Economic & Public Policy Research Group

Project Leaders
Alan Clayton-Matthews, Associate Professor Emeritus of Economics and Public Policy, Northeastern University

Thomas Peake
Branner Stewart

Project Staff
Lily Harris

Unit Director
Mark Melnik, Director of Economic & Public Policy Research

Established in 1971, the UMass Donahue Institute is a public service, research, and economic development arm of the University of Massachusetts. Our mission is to apply theory and innovation to solve real world challenges and enable our clients to achieve their goals and aspirations. We serve clients in the public, non-profit, and private sectors in the Commonwealth and throughout the nation and the world. For more information, www.donahue.umass.edu.

The Institute's Economic & Public Policy Research (EPPR) group is a leading provider of applied research, helping clients make more informed decisions about strategic economic and public policy issues.

EPPR produces in-depth economic impact and industry studies that help clients build credibility, gain visibility, educate constituents, and plan economic development initiatives. EPPR is known for providing unbiased economic analysis on state-level economic policy issues in Massachusetts and beyond and has completed a number of industry studies on IT, defense industries, telecommunications, health care, and transportation. Their trademark publication is called *MassBenchmarks*, an economic journal that presents timely information concerning the performance of and prospects for the Massachusetts economy, including economic analyses of key industries that make up the economic base of the state.

UMassAmherst | Donahue Institute
Economic and Public Policy Research

Table of Contents

List of Figures

Summary

Baseline payroll employment by industry super sector for the 13 Regional Planning Agency (RPA) regions in 2020 and projections for 2030, 2040, and 2050 are presented in Table 1[1]. The baseline 2020 is ES-202 payroll employment. The jobs projections are consistent with population projections by age, sex, and RPA, and with labor force projections by age, sex, educational attainment, and RPA. They assume that the state's economy is at full employment by 2030 with a NAIRU (non-accelerating inflation rate of unemployment) unemployment rate of 3.6 percent.

[1] Due to their size, the tables referred to in this document could not be included in the document itself. Please refer to the associated Excel workbook "Massachusetts State and Regional Employment Projections Tables".

Components of Methodology

The methodology involves several components related to forming the projections, including the various concepts of employment that relate population and labor force to payroll jobs, sources of data and information for the projections, and methods and models that use the data to form the projections.

Employment Concepts

Labor force (resident) employment

Population, and labor force and its components: employment and unemployment, are residence-based and in units of persons, that is, persons and labor force members are counted where they live. A working-age person (16 years or older) is either in the labor force, i.e., participating in the labor force, or not. The measure of the labor force is the number of persons who are participating. Participants are either employed (worked last week for pay) or unemployed. The measure of employment is the number of persons employed, the measure of unemployment is the number of persons unemployed, and the unemployment rate is the number of unemployed persons divided by the number of persons in the labor force. We refer to these below as labor force employment concepts or residence-based employment, often interchangeably.

Payroll jobs by place of work

On the other hand, the final projections of employment are a different concept. They are the number of payroll jobs by place (region) of work. This concept, as measured by the ES-202 series, is used for final projections because they are based on a full count from administrative records of persons on payrolls of employers who are required to report to the unemployment insurance system and tax authorities. The ES-202 covers the vast majority of workers, and is a reliable measure of this concept, not subject to the sampling variation of labor force estimates, which are based on samples. This concept differs from the labor force concept of employment in several ways:

- It is not a count of persons, but of jobs. A person may work for more than one employer (a multiple job holder), and so may be counted more than once in this concept.
- Jobs are counted at the location (region) of the establishment rather than where the worker lives.
- Persons who live outside of Massachusetts but work in Massachusetts are not part of Massachusetts residence employment; and conversely, persons who live in Massachusetts but work out of state are not part the Massachusetts payroll employment.
- It excludes jobs that are not on an employer's payroll, that is, it excludes self-employed jobs.

Accounting relationships between employment concepts

Measures of the same concept can be compared and manipulated. Those of different concepts cannot.[2] However, the employment projections need to use both concepts. The labor force concept ensures consistency between the residence-based concepts of population, labor, force, employment, and unemployment targets by region. At the same time payroll jobs by place of work projections must be consistent with the ES-202 and equal to the ES-202 by industry and region in the baseline year. A framework is needed to be able to transition between one concept and another.

This is approximately implemented by an accounting framework that can transform one concept to another by adjusting for the differences in the two concepts noted above. Each step transforms the units of one concept into the units of another by series of linear factors derived from data measures on each concept. Each factor is estimated for each detailed industry at the state level. These are defined in the call-out box "Jobs and Employment Accounting Relationships", and the estimates of the factors by industry are presented in Table 2. The factors are:

- α (alpha): The ratio of jobs by place of residence to jobs by place of work. Alpha is estimated from a LODES (LEHD Origin-Destination Employment Statistics where "LEHD" stands for the U.S. Census Bureau's Longitudinal Employer-Household Dynamics) analysis of net commuting patterns into Massachusetts from other states, using baseline resident employment from the PUMS (the American Community Survey's Public Use Microdata Sample). At the state level, there is little variation in these parameters by industry because there are few LODES origin/destination industry sectors. Part of the small state-level variation is due to the different geographical distributions of employment by industry. These parameters are all less than one, reflecting net in-commuting into Massachusetts. The alphas in Table 2 are used for state level resident employment estimates and parameters only. The projections by industry and region use a LODES commuting model in which alpha varies significantly by region as well as industry.
- δ (delta): The ratio of jobs that are held by multiple job holders. Delta is estimated from the U.S. Census Bureau'sCurrent Population Survey (CPS) which asks about dual job holding. This parameter varies significantly by industry. $(1 + \delta)$ is the ratio of jobs to employment, $(1 + \delta) \geq 1$.
- γ (gamma): The ratio of self-employment to payroll employment. Gamma is also estimated from the CPS, using the class of worker (employee vs. self-employed status), and varies significantly by industry. Note that $1 + \gamma$ is the ratio of employment to payroll employment, so $1 + \gamma \geq 1$. γ's are estimated separated for 2019 and 2020. The 2020

[2] For example, the labor force concept of the unemployment rate is a valid measure, since its components are parts of the same concept: labor force = employment + unemployment. However, an unemployment rate calculated using payroll jobs by place of work divided by labor force (persons by place of residence) cannot and makes no sense. This could result in a negative unemployment rate calculation!

value is applied in baseline (2020) estimates, and the 2019 value is applied in 2030, 2040, and 2050 projections. The reason for this choice is that 2020 is the year of COVID recession. In any case, these parameters are very similar in both years.

These parameters involve assumptions that are approximations to the truth, so there can be slippage in transforming from one employment concept to another to the extent the assumptions are violated. It is assumed that δ and γ do not vary over time or region. Limited analysis seems to confirm that they are reasonable stable over time, and it is plausible that the variation is closely related to industry characteristics rather than region of residence. Their linearity means that they can be combined in any order to convert from one employment concept to another. For example, in developing the state level residence employment baseline and targets from the 2019 and 2020 ES-202 actual data, a transformation needs to be made from payroll jobs by place of work by industry to labor force (resident) employment. This is accomplished by multiplying payroll jobs by place of work by alpha, which gives payroll jobs by residence (that is, transforming the number of payroll jobs in Massachusetts to the number of payroll jobs held by residents of Massachusetts), then dividing by the quantity 1+delta, which transforms the number of payroll jobs held by residents of Massachusetts to the number of payroll resident workers in Massachusetts, then multiplying by 1+gamma, which transforms the number of payroll resident workers in Massachusetts to the number of resident workers in Massachusetts by adding resident self-employment.

Jobs and Employment

The accounting relationships described below apply to each of the 60 detailed industrial sectors at the state (Massachusetts) level. They are used to transform employment from one concept to another.

Definitions

$\boldsymbol{PJ_w}$: Payroll jobs by place of work. Source: ES 202

$\boldsymbol{PJ_r}$: Payroll jobs by place of residence.

$\boldsymbol{PE_r}$: Resident payroll employment.

$\boldsymbol{SE_r}$: Resident self-employment. Source: CPS

$\boldsymbol{E_r}$: Resident employment (labor force employment), where $E_r = PE_r + SE_r$.

$\boldsymbol{\alpha}$: Ratio of resident workers in a region to place of work employment in a region. Source: LODES.

$\boldsymbol{\delta}$: Parameter derived from the CPS. The proportion of jobs that are held by multiple jobholders.
$(1+\delta)$ is the ratio of jobs to employment.

$\boldsymbol{\gamma}$: Parameter, ratio of self-employment to payroll employment, where $\gamma = \frac{SE_r}{PE_r}$.

Accounting Relationships

$PJ_r = \alpha PJ_w$.

$PE_r = \frac{PJ_r}{1+\delta}$.

$PE_r = \frac{E_r}{1+\gamma}$.

Sources of Data and Information

The projections are based on data from several public sources:

- ES-202 payroll employment, a census of establishments that report to state and federal governments that includes counts of the number of employees on the payroll. The projections use two versions of the annual files that report average monthly payroll employment:
    - The state-level ES-202 by detailed industry to provide the baseline state payroll employment.
    - An RPA-level ES-202 by industry supersector constructed by MAPC. MAPC built this version by aggregating municipal-level payroll employment, imputing the value of censored cells at the municipal level by allocating state total payroll counts to the censored cells preserving known municipal and state totals. The allocation was done by detailed industry sector. This version was used to construct the baseline payroll employment by RPA and industry supersector.
- Bureau of Labor Statistics (BLS) employment projections 2020-2030 (https://doi.org/10.21916/mlr.2021.20); and 2019-2029 (https://doi.org/10.21916/mlr.2020.21). These sources are used to form the state-level projections for resident employment by detailed industry. The BLS provides 10-year projections of population, labor force, output, and employment by detailed industry and occupation at the national level.
- U.S. Census Bureau American Community Survey (ACS) Public Use Microdata Sample (PUMS), 2015-2019. This source is used to distribute state-level resident employment projections by detailed industry to the RPAs.
- U.S. Census Bureau ACS Summary File (ACSSF) tables, 2015-2019. These are used to identify RPA geographies on the PUMS.
- BLS Current Population Surveys (Basic version), January 2019-December 2021. This monthly survey is used to estimate baseline state-level self-employment by detailed industry, and to estimate the delta and gamma parameters used for transformations between different employment concepts (see above section on Employment Concepts).
- U.S. Census Bureau LEHD Origin-Destination Employment Statistics (LODES), 2015-2019. This dataset is used to distribute employment by place of residence (RPA region)

to employment by place of work (RPA region and out-of-state), and to distribute non-Massachusetts resident employment to RPAs.

Tools and Models

Two software applications used with the PUMS dataset and a model describing the procedure for estimating LODES-based commuting parameters are used to generate the regional distribution of employment projections.

- Least squares reweighting, aka the "reweighter", generates synthetic populations from the PUMS by factoring the household and person weights on the PUMS to hit user-specified targets for the synthetic population to be generated. The targets set by the user can be any function of the variables present on the PUMS, for example, a target population by sex and age cohort. The output of the reweighter is a set of household and person weights that can be merged onto the PUMS, providing a new, synthetic PUMS that is identical to the input PUMS except for having a different set of weights. Tabulation of the synthetic PUMS using the same function as that for the targets will provide results that "hit" the targets, usually with great precision. The weight factor for the person weights in a given household is the same as the weight factor for the household weight. See the call-out box "How the Reweighter Works". For a comparison of the reweighter to iterative proportional fitting (IPF) and other similar synthetic methods see (Clayton-Matthews, 2019).
- PUMA splitting, aka, the "splitter" is used to "identify" the geography codes on the PUMS that were censored by the Census Bureau. The splitter does not correctly assign a significant proportion of geographies but does assign geography codes such that tabulations of the PUMS with the assigned geography codes closely approximate the ACS SF tables for that geography, providing an "as if" identification. The splitter does not change the input PUMS in any way. It simply provides an additional geography code that can be merged with the PUMS. In the context of this project, it is used to define RPA regions on the PUMS. An RPA is a set of municipalities (level 60 geographies in the Census Bureau's classification). If a PUMA crosses one or more RPA boundary, the splitter can be used to divide the PUMA into smaller geographies such that each geography lies in only one RPA. See the call-out box "How the PUMA Splitter Works".
- The LODES commuting model is an accounting guide for using the LODES origin/destination data to calculate the commuting parameters used to transform place of residence employment concepts into place of work concepts. The LODES data observations used in this analysis are counts of the number of persons who live in region r_1, work in region r_2, and work in industry i. LODES provides geographies down to the block group level, but only provides three industry sectors. The accounting guide indicates how the commuting parameters are constructed from the origin/destination data. These parameters are calculated for the 2015-2019 period and are assumed to be constant over time. See the call-out box "How LODES Works" and the call-out box "LODES Commuting Model".

How the Reweighter Works

The reweighter is used to generate a synthetic population from the PUMS that, upon tabulation, reproduces one or more tables of targets. For example, suppose you wanted to analyze the effect of changes in state-level Massachusetts employment by industry and occupation on unemployment, wage and salary incomes, location of residence, demand for housing, etc. Form a table or tables with the target employment, that is, the current employment by industry and occupation plus the changes in each to form the target. This could, for example, be a table of 50 industries by 15 occupations. If this takes place over time and other factors change, for example, population, you could also form tables of population estimates by age, sex, etc. to reflect the view of population at the target date. The only requirement for the target tables is that the PUMS could be used to calculate the cell contents of the target tables.

The reweighter then calculates a new set of household and person weights, that, when used to tabulate the PUMS, gives tabulated tables that match the target tables. For each household, the household weight and the person weight change in the same proportion, so that the household composition does not change. What the weights *do* change, however, is the number of such households. The reweighter does not change any of the variables on the PUMS. It simply provides an alternative set of household and person weights that “hit” the targets. The PUMS with the alternative set of weights is called a synthetic population that could be used for analysis in the same way you would use the PUMS. For example, you could analyze the change in employment and incomes by region, sex, etc.

Here is how it works. It tabulates the PUMS to create tables that correspond to the target tables, beginning with the original household and person weights on the PUMS. Of course, the original set of weights will not produce tables that match the targets. Each cell in each PUMS-tabulated table is compared to the corresponding cell in the target table, and the squared difference between the two cells is formed. This is done for each cell in all the tables, and these squared differences are all added up, forming the so-called sum of squares or sum of squared deviations. Here this sum of squares is called the value of the objective function (OF). If the tabulated tables matched the target tables exactly, each deviation would be zero and so the value of the OF would be zero. The weights that give this result are the desired weights. Any deviation from an exact match results in a positive value for the value of the OF, with higher values indicating a worse match.

(See the “Flow Logic for Reweighting” diagram.)

On the other hand, the value of the OF increases, the nudge in the weights worsen the fit, and the new set of weights for that household is rejected. In this case a small proportional *decrease* in the household's weights are tried instead. If *that* change lowers the value of the OF (improves the fit), then it is accepted as the household's new set of weights. If neither a small positive change nor a small negative change lowers the value of the OF, no change is made in that household's weights.

The algorithm proceeds to the next household, and the process is repeated for this second household. This continues, household by household; and then after all households have been processed, begins another round looping through all the households. Each time a set of weights has been changed, the value of the OF falls slightly. This continues until the OF is zero, when the PUMS tabulated tables and the target tables match exactly (or close enough, as specified by the user).

Initially, the tables have a bad match and the value of the OF is very large. The algorithm takes the first household in the PUMS and increases the household weight and each person weight by a small proportion, say 1%. It then tabulates the entire PUMS with these new weights to form tables that correspond to the target tables and calculates the value of the OF again.[3] If the value of the OF decreases, the small increase in the household's weights has improved the fit and the new set of weights is accepted for that household. If, on the other hand, the value of the OF increases, the nudge in the weights worsen the fit, and the new set of weights for that household is rejected. In this case a small proportional *decrease* in the household's weights are tried instead. If *that* change lowers the value of the OF (improves the fit), then it is accepted as the household's new set of weights. If neither a small positive change nor a small negative change lowers the value of the OF, no change is made in that household's weights.

The algorithm proceeds to the next household, and the process is repeated for this second household. This continues, household by household; and then after all households have been processed, begins another round looping through all the households. Each time a set of weights has been changed, the value of the OF falls slightly. This continues until the OF is zero, when the PUMS tabulated tables and the target tables match exactly (or close enough, as specified by the user).

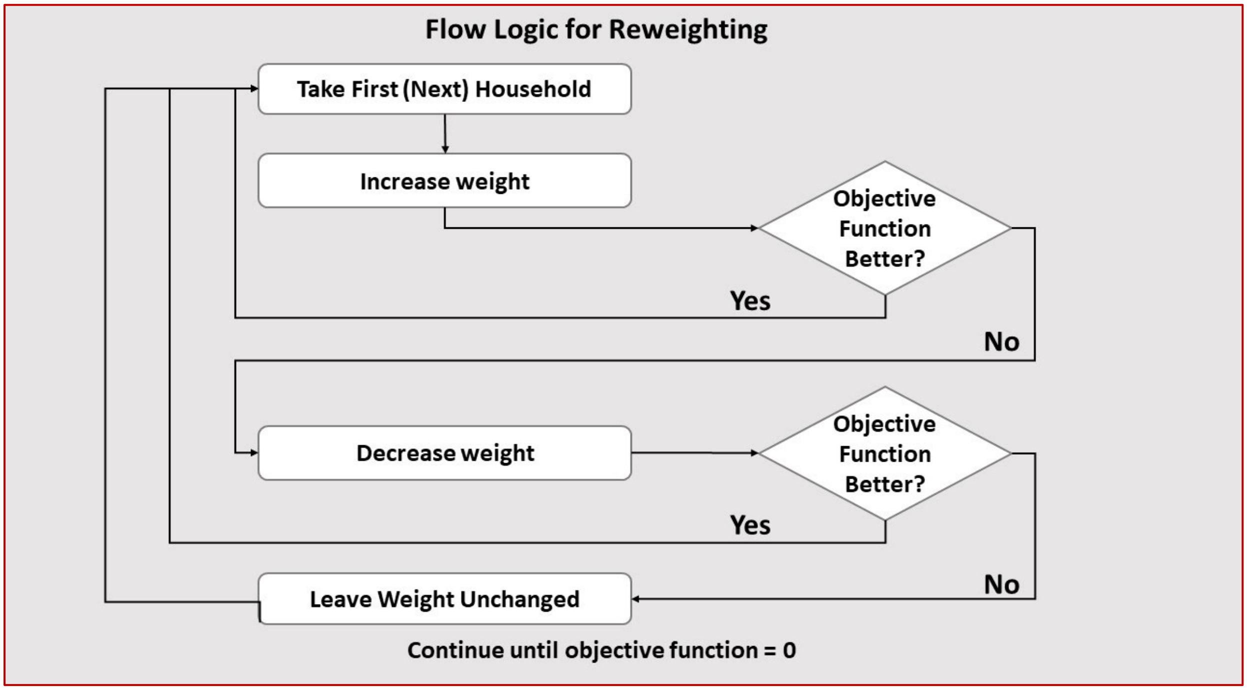
Although this algorithm may seem random – doesn't it depend on the order of the households? – It yields a unique solution. In fact, it reaches the solution faster if the households are shuffled after each round through the PUMS, which is what the algorithm does.[4]

---

[3] Conceptually, it redoes the tabulations "from scratch", but actually the value of the OF can be updated without having to retabulate after each trial set of weights.

[4] The user can set the proportional nudge to whatever value they like. The uniqueness property is in the limit as the proportional change approaches zero, but execution time is longer the smaller the nudge, for example, 0.1% instead of 1% would increase the execution time 10 times. The simple 1% nudge works well and produces a set of weights very close to the unique limit.

Figure 1: Flow Logic for Reweighting



How the PUMA Splitter Works

The Census Bureau's American Community Survey (ACS) is a representative sample of U.S. households and the persons who live in those households, sampled at a rate of about 1.3% per year. From the detailed demographic survey of the respondents, the Census Bureau produces two products:

A series of a few hundred tables, the ACS Summary File (ACSSF), that are tabulations and cross-tabulations of key variables from the survey.

The Public Use Microdata Sample (PUMS), which includes, for each household and the persons who live in that household, the complete set of survey responses. For every 13 households in the ACS, 10 are selected for inclusion in the PUMS product.

The ACSSF product produces the tables for a wide spectrum of geographies, from the nation down to the block group level for their 5-year products. For the PUMS product, however, only a few geography codes are included: state, division (e.g., New England), and Public Use Micro Area (PUMA). A PUMA is a geographic area with a population of 100,000 or more persons.[5]

Suppose you wanted to use the PUMS to study the city of Everett, Massachusetts. Everett is part of a PUMA that also includes Somerville, the Somerville/Everett PUMA. If you could identify which sample households were from Everett, and which from Somerville, then you could limit your analysis to the Everett sample households. This is what the splitter allows you to do.

Here is how it works. Conceptually, tabulations and cross-tabulations of the Everett sample from the PUMS would match the corresponding tables in the ACSSF, and of course, those of the Somerville sample would match the ACSSF tables for Somerville. The splitter tries different combinations of assignments of the PUMS sample households to Everett and Somerville until the PUMS tabulations match the ACSSF tables as closely as possible.

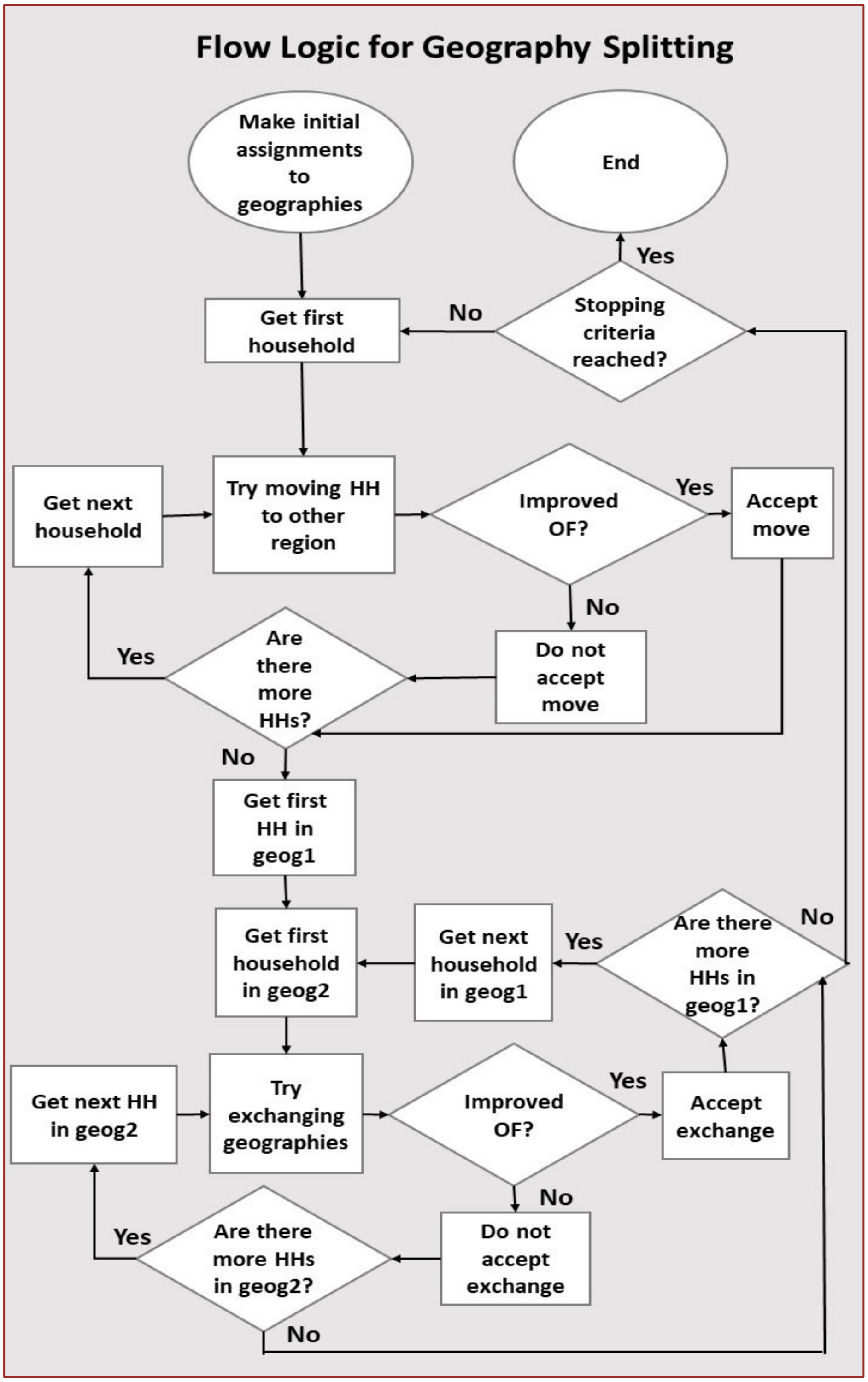
[5] In the mid 1990's Williamson, Birkin, and Rees (1998) faced the exact same problem of censored geographies in the microdata file from the 1991 United Kingdom census. Their approach was similar, using the same objective function but operationalized in a different manner.

To assess how well the proposed assignment reproduces the ACSSF tables, it compares each cell in a table tabulated from the PUMS to the corresponding cell in the ACSSF table. Each cell difference is squared, and these squared cell differences are then added together into a number called the value of the objective function (OF). If the two sets of tables match exactly, each difference is zero and the OF equals zero. Any deviation from an exact match results in a positive value for the OF, with higher values indicating a worse fit. The splitting algorithm continues until it can no longer improve, i.e., lower, the value of the OF. For various reasons, an exact match is not possible[6], but the final assignment of each household to either Everett or Somerville produces tabulated tables from the PUMS that closely match the corresponding tables from the ACSSF. This could be called "as if" identification of the Everett and Somerville geography codes. Experience shows that the more ACSSF tables used, the better the fit. A test of the procedure under ideal conditions and with a known result – two PUMAs combined to form a super-PUMA – correctly placed 90% of the households into the correct geography.[7] For the other 10% of households, the characteristics of the misplaced households closely matched those of the "correct" households, and so analysis using the proposed geographies posed no problem. This is what is meant by "as if" identification. See the "Flow Logic for Splitting Geographies[7] diagram for the splitting algorithm.

[6] The PUMS, being a (large) subsample of the ACSSF, would only by small chance reproduce tables that exactly match their ACSSF counterparts. Also, the two products are separately weighted, so even at the PUMA level, only certain key variables would sum to equality. Also, the number of possible combinations makes finding the correct one infeasible to compute, so it would be virtually impossible to find it even if an exact match existed. It is also possible for two different combinations to give the same result, so then, which is the correct one? Finally, the splitting algorithm appears to stop too soon, hitting a local minimum of the OF, rather than the global minimum. There are ways to improve the algorithm.

[7] The ideal situation is one in which the PUMS tabulations could exactly match the ACSSF tables and the two sub-PUMA regions are heterogeneous, that is, the more different the households are, the easier it is to assign them to the correct geography. (By the way, this is good reason to use the splitter if you think the sub-PUMA of interest is different than the rest of the PUMA.) The test involved forming a super PUMA from two very different PUMAS – the Mattapan/Roxbury Boston PUMA and the PUMA that included Dover, Needham, Sherborn, Wayland, Wellesley, Weston, and Westwood. The "ACSSF" tables for the two sub-super PUMAs were tabulations of the two respective PUMAs from the PUMS. The test was designed to see the possible potential for the algorithm.

Figure 2: Flow Logic for Geography Splitting



How LODES Works

The LODES (LEHD Origin-Destination Employment Statistics) datasets are produced by the Longitudinal Employer-Household Dynamics (LEHD) program, which is part of the Center for Economic Studies at the US Census Bureau. The datasets themselves are created using Unemployment Insurance earnings data and Quarterly Census of Employment and Wages (QCEW) data, which states may elect to share with the Census Bureau through the Local Employment Dynamics (LED) Partnership. LODES data is available at the census block level. In Massachusetts, the most recent LODES data available is for 2019. This study uses an average in LODES datasets between 2015 and 2019.

LODES data is presented in three files: the Origin-Destination (OD) file, the Residence Area Characteristics (RAC) file, and the Workplace Area Characteristics (WAC) file. This analysis used only the OD file. LODES OD data is arranged by origin-destination sets, with each observation being the combination of a single worker origin (place of residence) and destination (place of work). For each origin-destination combination, LODES data gives the total number of workers whose own place of residence and place of work align with that combination. In addition to the total number, LODES data also provides detail on those workers by broad age, income, and industry data. This analysis did not make use of the age or income data, but did use the industry data, and aligned LODES' three broad groups of industry sectors (Goods Producing, Trade, Transportation, and Utilities, and All Other Services) with ES-202 data.

LODES datasets are state-specific, with most, but not all, states participating in the program at this time. In addition to the main files, which contain origin-destination data for all workers living and working in that state, LODES also creates auxiliary files, which contain origin-destination data for all workers with workplaces in a given state but residences outside of that state. In order to capture cross-border commuting for this analysis, the auxiliary files for Massachusetts, the other five New England states, New York, and New Jersey were analyzed. While there are a small number of workers in Massachusetts with a state of origin outside of this region, as well as Massachusetts residents with workplaces outside of this region, these are likely unique working arrangements, which do not reflect regular commuting patterns. For the purposes of creating an employment projection for MassDOT, the decision was made to exclude these cases.

LODES Commuting Model

Notation:

w : the number of workers, employment, or jobs.

i : indexes the LODES super sector industries, i=1:3.

p : indexes region segments, p=1:61. The segments consist of 37 whole PUMAs, Boston, and 10 PUMAs that are split into 23 sub-PUMAs.

r : indexes RPAs, r=1:13.

o : indicates the out-of-state region.

P : indexes regions, P={p, o}, P=1:62.

R : indexes regions, R={r, o}, R=1:14.

$\bar{r}$: a bar indicates the complement, e.g., r¯indicates all regions excluding r.

w(i) is the number of workers in industry i.
Superscripts indicate region of residence, subscripts indicate region (place) of work. Resident employment in a region is the number of workers who live in that region, in contrast to place of work employment in a region, which is the number of workers who work in the region.

Origin/Destination Matrix

The LODES database is used to estimate an origin/destination matrix for the baseline and each target year, where each cell, $w(i)_R^P$, is the number of workers in industry i who live in region P and work in region R. In all, this matrix has (62 x 14 -1) x 3 x 4 years, or 10,404 cells ($w(i)_o^o$ is not included and appears as zero in the matrix). The matrix is in Appendix Table 1. This matrix is used to form several measures, including some useful ratios that can be used as parameters to characterize the commuting patterns of regions. These ratio parameters are constructed from the 2015-2019 LODES origin/destination dataset and do not vary by projection year.

Out-commuting to other regions

One useful ratio is the number of out-commuters from a region per resident employment worker in the region. This ratio is: $Ratio(i)_R^p = \frac{w(i)_R^p}{w(i)^p}$. Note that the sum of the ratios for each segment/industry combination equals 1: $\sum_R Ratio(i)_R^p = 1$.

In-commuting from other regions

Another useful ratio is the number of in-commuters into a region per resident employment worker in the region. This ratio is: $Ratio(i)_r^R = \frac{w(i)_r^R}{w(i)^r}$.

Alpha Parameters

The alpha parameter for a region is the ratio of its resident employment to its place of work employment. This ratio is: $\alpha(i,r) = \frac{w(i)^r}{w(i)_r} = \frac{w(i)^r}{w(i)_r^{\bar{r}}+w(i)_r^r}$. This ratio is given a parameter name because of its usefulness in transforming one employment concept into another.

Employment Decomposition

The origin/destination matrix is also used to decompose place of work employment into its components.

$$w(i)_r = w(i)^r_{\square} - w(i)^r_{\bar{r}} + w(i)^{\bar{r}}_r = w(i)^r_r + w(i)^{\bar{r}}_r$$

The number of workers who work in region r is equal to the number of workers who live in region r minus the number of workers who commute out of region r to other regions plus the number of workers who live in other regions and commute into region r. Alternatively, it is the number of workers who both live and work in region r plus the number who commute into region r from other regions.

Project Steps

Step 1: State level resident employment projections

BLS 10-year employment projections, shift/share methodology

State level employment projections for 2020-2030 are derived from the 2020-2030 and 2019-2029 long-term employment projections from the BLS. The national BLS employment projections are used because of their superior quality. They form a consistent set of population, labor force, output, and employment projections. Like the SEPC process, population and changes in labor force participation are the main drivers of labor force growth. Projections for GDP, output, productivity, and employment use a well-established private sector macroeconomic model.

Their projections are intended to capture structural changes like consumer preferences and technology that affect what is produced and how it is produced over the 10-year horizon. They do not incorporate cyclical factors in their projections – although the position of the economy in the business cycle does, of course, determine baseline measures of many components like employment. They assume the economy is at full employment with a NAIRU unemployment rate at the 10-year horizon.

To derive the employment projections for Massachusetts from the national estimates, the well-established shift/share methodology is used. This is an appropriate methodology when national projections are at a detailed industry level. Structural factors that affect demand, output, and employment in specific industries are most often common across regions. For example, a shift in consumer demand from brick-and-mortar retail stores to e-commerce affects retail vs. transportation employment in all states. Technology that affects how goods are produced and services provided are common to a particular industry regardless of where it is located. This reflects the “shift” in shift/share. Industry concentration differs across regions, for example, IT and life sciences in Massachusetts vs. mining in the West. So aggregate employment growth summing across all industry sectors can differ depending on the relative concentration of industries across regions. This reflects the “share” in shift/share.

The BLS’s model-based employment projections include hundreds of detailed industries, which are aggregated to 60 detailed industry sectors for this project. This number gives enough of a variation in concentration rates by sector to take advantage of the shift/share approach while ensuring sufficient sample sizes on the CPS and PUMS for reliable analysis. In the final analysis in part 2 of Step 4, these 60 industries are aggregated to 10 super sectors (Table 3).

Isolating the COVID cycle effects from the 2030 projections

The BLS' methodology focuses on modeling the effects of structural change, so for their 2020-2030 projection structural change estimates were applied to 2019 employment levels. The economy in 2019 was at full employment, so their projections jumped off from 2019, and structural change growth rates from 2019 to 2030 applied to 2019 yielded the 2030 full employment projection. This conveniently allowed them to break down the 2020-2030 change in employment into two components: that due to recovery from COVID, and that due to structural change.

The same strategy is used for Massachusetts. Preliminary estimates of employment by detailed industry for 2030 use the individual sector structural growth rates from 2019 to 2030 given by the BLS 2019-2029 and 2020-2030 projections.[8]

Forming baseline resident labor force employment by industry, 2019 and 2020

The baseline payroll jobs by place of work are simply the ES-202 for 2020 by detailed industry. However, to obtain employment projections for 2030, 2040, and 2050 that are consistent with population and labor force projections in those years *and* that are consistent with the ES-202 in 2020 and 2019, the labor force resident employment concept must be constructed for 2019 and 2020. The latter year is the baseline year for the project – which reflects the COVID recession, and 2019 is the jumping-off year for structural change growth. This involves transforming payroll jobs by place of work into payroll resident employment and adding in self-employment.

This is accomplished in the following steps for each of the 60 detailed industries in 2019 and 2020 (see the section on employment concepts above):

1. Factor the ES-202 jobs count by the alpha factor (Table 2). This gives payroll jobs by residence, i.e., it converts the number of payroll jobs in Massachusetts to the number of payroll jobs held by residents of Massachusetts.
2. Divide the result by 1+delta (Table 2), which transforms the number of payroll jobs held by Massachusetts residents to the number of payroll resident workers in Massachusetts.
3. Add self-employment (a resident labor force concept) from the CPS.
4. This gives resident employment.

[8] Note that directly applying the 2020-2030 growth rates from the BLS 2020-2030 projections and ignoring 2019 would not have been the right thing to do. The COVID recession in Massachusetts was deeper, and the proportional loss in employment varied by Massachusetts versus U.S. industry.

Forming statewide total employment targets for 2030, 2040, and 2050

State resident employment in 2030 must be consistent with both state population and state labor force projections for 2030. Also, like the BLS projections, employment should be "full employment" consistent with the Massachusetts NAIRU in 2030.[9]

The BLS estimate of the NAIRU unemployment rate for the U.S. in 2030 is 4.3%. Given the demographic differences between Massachusetts and the U.S., the NAIRU for Massachusetts has for a long time been about 0.7% lower than that for the U.S. These rates are based on the Current Population Surveys (the headline U-3 unemployment rate). The labor force targets in this project are from the PUMS rather than from the CPS, and unemployment rates and labor force on the PUMS tends to be about one percentage point higher than on the CPS.[10] Therefore the NAIRU unemployment rate in 2030 for Massachusetts consistent with the PUMS is 4.6% = 4.3% - 0.7% + 1.0%.

The labor force targets from MAPC exclude the labor force in noninstitutional group quarters (GQ here, for short), but the population targets include GQ, and the ES-202 includes jobs held by persons living in GQ, so these are taken into consideration in forming target employment. GQ labor force participants accounted for 2.14% of the total labor force in the 2015-2019 PUMS. This factor was applied to the 2020 baseline, and 2030, 2040, and 2050 state labor force targets from MAPC to form the total labor force needed to calculate employment targets. (employment = [1-NAIRU] x labor force).[11] [12]

Given the population and labor force targets for 2020 and 2030, the corresponding resident employment target growth rate from 2020-2030 is 11.16%. This is substantially higher than the labor force target growth of 1.56% over the same period. The difference represents the recovery from the COVID recession and the associated fall in the unemployment rate.

[9] It is appropriate to follow the BLS convention of full employment in the projection year. The BLS does this so that changes in 10-year projections (from year-to year, say 2019-2029 vs. 2020-2030, or 2030 vs. 2020) reflect changes in outlook due to structural change. This is useful for policy analysis, for example, in designing workforce and training programs. Moreover, predicting at what point in the business cycle the economy is in 10 years is a fruitless task. Similar reasons apply to this project. For transportation planning, it seems appropriate that capacity needs should be calibrated to full employment.

[10] The reason may be that the CPS asks a series of questions to determine if a person is unemployed rather than out of the labor force. These are not asked in the American Community Survey (the source for PUMS), and so some persons who are actually out of the labor force by CPS standards may report themselves as being unemployed (and therefore in the labor force) on the ACS.

[11] GQ employment accounted for 1.96% of the total employment in the 2015-2019 PUMS. This proportion was applied in 2020, 2030, 2040, and 2050 to set targets for statewide GQ employment in the next (reweighter) step.

[12] Statewide estimates of statewide GQ labor force and GQ employment were used as additional targets in addition to the MAPC household labor force targets in the next (reweighter) step.

For 2030 to 2040 and 2040 to 2050, the NAIRU is assumed to remain the same, at 4.6%, so aggregate employment and labor force grow at the same rates. This is -.30% for 2030-2040 and 1.05% for 2040-2050.

Forming the employment by industry projections 2030, 2040, and 2050

Preliminary resident employment by detailed industry for 2030 was obtained by applying the industry-specific sector structural growth rates for the U.S., 2019 to 2030, to the Massachusetts 2019 baseline employment by detailed industry. This resulted in a 12.52% increase in employment over the decade, only slightly higher than the target growth rate. Final resident employment targets by detailed industry for 2030 were obtained by scaling the structural growth rates proportionately down to hit the target.

Resident employment projections for 2040 and 2050 were obtained in a similar manner. Inspection of the results for 2030 for plausibility and lack of knowledge about changes in relative structural growth rates after 2030 suggested that applying the 2019 to 2030 structural rates of growth to the 2040 and 2050 was reasonable. In each case, the structural growth rates were scaled downward to hit the targets.

The 2019 and 2020 baseline, and 2030, 2040, and 2050 projections for state resident employment by detailed industry, along with the corresponding U.S. structural growth rates, 2019 to 2030, are presented in Table 4.

Step 2: Distribute resident employment projections to the RPAs

Using the PUMS for the regional distribution of employment by industry

The American Community Survey's (ACS) Public Use Microdata Sample (PUMS) provides the best way to distribute projected resident employment by region. Households and the persons who live in those households are drawn from the ACS in a systematic way to maintain the geographic representation of the ACS down to the tract level.[13] The annual ACS is approximately a 1.3 percent sample of U.S. households, while the PUMS is a 1 percent sample, so about 10 out of every 13 ACS households are selected for the PUMS.[14] Because the PUMS is a smaller sample than the ACS, its sampling error is larger, but not by much.[15]

[13] U.S. Census Bureau, Public Use Microdata Sample (PUMS): Accuracy of the Data (2015-2019). https://www2.census.gov/programs-surveys/acs/tech_docs/pums/accuracy/2015_2019AccuracyPUMS.pdf

[14] The sampling interval depends on the number of households who responded to the ACS and so can vary from year to year and from region to region. The 1.3 figure cited here may not be correct for the 2015-19 PUMS; this estimate is subject to confirmation. According to Matthew Brault, a former survey statistician from U.S. Census Bureau, now a Senior Research Scientist with NORC and a member of the ACS data users group, the sampling interval is about 1.25 for the 2021 PUMS.

[15] Since the sampling error varies inversely with the square root of the sample size, the sampling error on the PUMS would be about 14% higher than that of the ACS for a sampling interval of 1.3.

The methodology assumes that the geographic distribution of workers by industry evolves slowly enough over time so that the existing distribution is a good estimate of the future distribution for the period of this study. This in turn embodies two other assumptions: that the distribution of establishments by industry and commuting patterns of workers also evolve slowly.

Side effect: the PUMS can provide a synthetic population

Geographic representation is not the only advantage offered by the PUMS. The dataset provides a rich demographic set of information that allows the distribution of employment to be consistent with other aspects of the SEPC projections. Workers live in households, and the PUMS includes information for each person in the household: age, sex, race, educational attainment, labor force status, etc. To the extent that household formation and the demographic characteristics of household members are correlated with employment by industry, changes in the distribution of employment – i.e., the industrial composition of employment or the number of jobs in an industry – will be associated with changes in these household characteristics, given by the joint distribution of employment and household characteristics available on the PUMS. This correlation works both ways. As the population and labor force by geography changes, changes in the location of employment will be associated with these demographic changes. For example, if the labor force in one region is increasing while the labor force in another region is declining, the geographic distribution of employment between these two regions will shift from the shrinking region to the expanding region. These effects rely on the plausible assumption that these joint distributions of characteristics evolve slowly over time.

The upshot is that the PUMS provides a means to generate a synthetic population that responds to changes in population, labor force, and employment that is consistent with these projected targets. This feature may be desirable. See the call-out box "Coastal Community Synthetic Population".

Coastal Community Synthetic Population

Imagine a coastal community with a fishing industry, a resident population of artists, and a picturesque landscape that attracts retirees. The households in the town are fishing households, artist households, and retiree households. The three types of households have different characteristics on average, of course, in terms of family composition, size, age, presence of children, workers, income, etc.

The economy and demographics of the region are changing. The fishing industry is declining, and the population is aging as some families move out for better opportunities, and others move in, drawn by the pleasant environment.

These trends are captured in the model by targets that reflect declining employment in the fisheries, perhaps stable employment in leisure, hospitality, and the arts, and an aging population with more persons over 60 and fewer children and youth.

How does the reweighter incorporate these changes to the PUMS given by the targets? It does not change the persons in the sample households, replacing one type of person with another. Rather, it alters the household weights (and the person weights of the persons in the household in equal proportion), making more households look like retiree households, and fewer households look like fishing households.

This changes the region's household characteristics – household size, composition, etc., indirectly, without the need to develop separate targets for household characteristics. Are these the changes one would want? Probably. They are consistent with the changes given by the targets.

Identifying RPA regions on the PUMS

Although the American Community Survey – and for practical purposes, the PUMS – are geographically representative at the RPA level, the Census Bureau censors the geography code on the PUMS to the Public Use Microdata Area (PUMA) level, an area that includes at least 100,000 persons. To correctly distribute resident employment to the RPAs, the RPA geographic code for the PUMS is needed. These are given by the PUMA "splitter" tool. This tool uses ACSSF tables, which provide tabulations of key ACS variables at a fine level of geographic detail. RPAs are collections of municipalities, the summary level 60 geographic code in the ACS, so tables for each municipality are available. Each level 60 geography is contained within a single PUMA. They can be added together to form tables for any combination of municipalities that form part (or all) of a PUMA.

The splitter works by using the ACS tables as targets for tabulations of PUMS variables that reproduce the ACS tables from the PUMS microdata. To split a PUMA into two sub-PUMAs, the individual level 60 ACSSF tables are aggregated into the two desired sub-PUMA regions. The splitter assigns each household to one of the two desired sub-PUMA regions so that the tabulated PUMS tables hit the ACSSF targets. The targets cannot be hit exactly because the PUMS is a (large) subsample of the ACS, and the ACS and PUMS records are weighted separately. However, the splitter does achieve "as if" identification of the sub-PUMA region. The splitter does not change any variable or weight on the PUMS; it simply provides an additional geographic code for the PUMS, in this case the RPA geographic code. This RPA code can then be used to identify the RPA to which the household belongs. The RPA codes are given in Table 5.

For this project, 33 ACSSF tables were used for subject areas related to household tenure, household size, age, sex, race, Hispanic status, Hispanic origin, school enrollment, educational attainment, poverty status, earnings, household income, labor force status, industry and occupation, and group quarters population. The list of tables from the ACS is in Table 6. In all, 10 PUMAs (out of 52 PUMAs in all) were split into 23 sub-PUMAs, giving the PUMS 65 contiguous, non-overlapping regions (aka "segments"). A crosstabulation of PUMA and RPA regions is given in Table 7.

Distributing state level resident employment to the RPAs consistent with population and labor force targets

The distribution of state level resident employment to RPAs used the reweighter with population targets by age, sex and RPA from the Donahue Institute, household labor force targets by age, sex, educational attainment, and RPA from MAPC, state-wide resident employment targets by detailed industry, state-wide group quarters labor force, and state-wide group quarters employment.

The reweighter was applied to the 2015-2019 PUMS, producing four synthetic population PUMS datasets, for the baseline year 2020, and for the target years 2030, 2040, and 2050. In each case, the 2015-2019 PUMS was reweighted to the target year. Resident employment by detailed industry and RPA was calculated by tabulating the synthetic PUMS datasets and applying the δ parameters to convert resident employment to resident jobs. These jobs estimates for the baseline and projection years by detailed industry and RPA are in Table 8. The jobs estimates by detailed industry and segment region used in the next step are in Appendix Table 1.

Step 3: Distribute employment from place of residence to place of work, the LODES origin/destination dataset

The employment projections require the construction of two basic employment concepts: (1) employment by place of residence; and (2) employment by place of work. The former is needed to provide worker counts by residence consistent with population and labor force projections. The latter is needed to provide counts of payroll employment by place of work. These two concepts are related by commuting patterns from residence to place of work and vice versa. Together they provide a description of movement from one region to another useful for transportation analysis.

The origin/destination dataset matrix

The LODES origin/destination dataset, 2015-2019, was used to estimate the commuting patterns needed to distribute employment by place of residence to employment by place of work.[16] This dataset is based on administrative and other data that count the number of jobs and workers by their region of origin, region of destination, and super sector industry sector (3 sectors) at a detailed block group geography. The dataset is effectively a 3-dimensional matrix in which each cell is the count of workers in a super sector who live in the origin region and work in the destination region. These data are used to estimate parameters for three sets of key ratios by region of residence, region of work, and broad industry sector:

- The number of out-commuters from a region per resident worker in the region.
- The number of in-commuters into a region per resident worker in the region.
- The number of resident workers in a region per person working in the region, the so-called alpha parameter.[17]

These parameters were applied to the resident employment by segment and detailed industry from step 2 above to create a baseline and projected origin/destination matrix. In forming this matrix, resident employment was converted into jobs by detailed industry using the $1 + \delta$ factor (see Employment Concepts), where δ is the proportion of jobs in the industry that are secondary

[16] For a description of LODES, see the above sections on LODES in the "Tools and Models" section of this documentation, and the call-out boxes "How LODES Works" and "LODES Commuting Model".

[17] These are not independent parameters. Any one can be calculated from the other two, but each has their own usefulness in analysis.

jobs of dual job holders. For the purpose of estimating the key ratio parameters from LODES, jobs by detailed industry were aggregated into the three LODES super sectors:

1. Goods Producing
2. Trade, Transportation, and Utilities
3. All Other Services.

The industry detail present in Table Appendix 2 keeps this aggregation. For the final employment estimates in the next step, the LODES parameters were applied to the full set of 60 detailed industries. This assumes that within each origin/destination LODES super sector cell, the parameters are equal across the detailed industries, which is an approximation.

Regions and segments in the origin/destination matrix

For this analysis two levels of regional detail were used. For destinations (place of work), the regions are the 13 RPAs plus the out of state region. For origins (place of residence), a finer level of detail was employed to use more of the information present in LODES for better estimates of in- and out-commuting by region. In this report, they are referred to as region segments, or simply segments. Each segment is contained in one and only one RPA region, and the descriptor for that segment identifies the RPA region. In reading the origin/destination matrix, note that there are six types of segment descriptors:

1. The most common type is the PUMA level of geography for PUMAs that are contained within an RPA and form part of that RPA. The descriptor is “pt” followed by the RPA initialization followed by the Census PUMA geography code. For example, “pt MAPC 504” is PUMA 504 which is part of the MAPC RPA.
2. Part of a PUMA that was split so the PUMA part (segment) lies within a single RPA and forms part of that RPA. The descriptor is “pt” followed by the RPA name followed by the PUMA geography code that was split followed by the numeric code for the RPA (in initialization order) For example, “pt PVPC 200-12” is the part of PUMA 200 that lies in the PVPC RPA. The numeric code for PVPC is 12.
3. Part of a PUMA that was split so the PUMA part (segment) lies within a single RPA and forms the entire RPA. The absence of “pt” at the beginning indicates that the segment consists of the entire RPA. For example, “MVC 4800-7” is the MVC RPA that is part of PUMA 4800.
4. A PUMA that is congruent with an RPA. The descriptor is the RPA initialization followed by the PUMA geography code. For example, “BRPC 100” indicates that the BRPC RPA and PUMA 100 are the same geography.
5. Boston: “pt MAPC Boston”. Boston consists of 5 PUMAs and is part of the MAPC RPA.
6. “Out of State”: Indicates outside of Massachusetts.

Table 9 is the origin/destination matrix for the baseline and projected years aggregated across industry sectors. The last row of the matrix gives the number of out of state in-commuters into each RPA. In the base year, these amounted to 294,160 on an average working day. The last column gives the number of Massachusetts resident workers who commute to other states, 168,122 per day on average in the base year. Net commuting into Massachusetts from other states was 126,038 per day in the base year.[18]

Uses of the origin/destination matrix

The origin/destination matrix contains information useful for analyzing regions' characteristics of resident versus place of work employment and commuting patterns. Three examples are presented here.

1. ***The distribution of RPA place of work employment by region of residence***

Table 10 gives the distribution of RPA place of work employment for each of the segment's resident workforce estimated from the LODES parameters. For example, focusing on the Boston row of the table, 91.88% of Boston's resident workers work in the MAPC region, of which Boston is a part. The remaining 8% of Boston's resident workers work outside the MAPC region, with 1.47% working in the OCPC RPA and 1.74% working outside Massachusetts.

Focusing on the part of the OCPC RPA that lies within PUMA 4903 – Plymouth County (East), which includes Plymouth, Duxbury & Kingston – more resident workers work in the MAPC RPA, 42.32%, than in the OCPC RPA itself, 35.32%, reflecting the extent of commuting from these South Shore towns to the Boston area.

Several border segments have significant out-of-state commuting. For example, 7.13% of resident workers in BRPC RPA work out-of-state – presumably in the Albany area, 9.44% of resident workers in the FRCOG RPA work out-of-state – most likely in Vermont or New York. 10.12% of resident workers in the MVPC RPA work out-of-state – mostly in New Hampshire or Maine, 12.62% of PVPC resident workers work out-of-state – presumably in the Hartford area, and 23.66% of resident workers in the part of the SRPEDD that is PUMA 4200 – which includes Attleboro, North Attleborough, Plainville, Rehoboth, Seekonk, and Swansea – work out-of-state, most likely in neighboring Rhode Island.

2. ***Employment decomposition into resident employment, gross commuting, and place of work employment***

The origin/destination data can be used to decompose employment into its resident, place of work, and gross commuting flows. See Table 11. For example, the CMRPC had a baseline resident employment of 299,894. Of these resident workers, 142,806 commuted to work in other

[18] Note that in the discussion, which follows in this step, the metric is jobs not persons. "Persons" or "workers" or "commuters" are used here to avoid the awkwardness that would otherwise be needed, e.g., "Massachusetts resident jobs who commute to other states…". This means that the magnitudes presented here (but not ratios or proportions) would have to be deflated by the δ factor, about 4.5%, to transform them to persons, workers, commuters, etc. Also note that the examples here refer to 2020, a recession year. By 2030, all these magnitudes would be roughly 10% larger.

RPAs, while 101,387 commuted from other regions – including, perhaps, also from outside Massachusetts – to work in CMRPC. The number of people working in the CMRPC was 257,475. This is less than the number of resident workers by the net number of out-commuters (142,806-101,387). Contrast this with the MAPC RPA, in which there were more than 400,000 more persons working in the MAPC RPA than workers living in the MAPC RPA. For Massachusetts as a whole, 3.536 million workers lived in the state, while 3.662 million persons worked in the state in the baseline year. This reflects a net in-commuting of 126,000 workers.

3. *The summary commuting measure alpha, α*

A summary measure that can be used to compare regions' commuting characteristics is the ratio of the number of workers who live in a region to the number of persons who work in a region, which is called alpha here, and which is symbolized by the Greek letter α in much of this documentation. For so called "bedroom" communities, where a lot of people live but in which few work, alpha would be high, well above one. On the other hand, in business districts, in which a lot of people work but few live, alpha would be low, well below one. Alpha has the following characteristics:

- If $\alpha > 1$ out-commuting exceeds in-commuting, and vis versa for $\alpha < 1$.
- Wider variations of alpha from 1 among regions is associated with higher and often uneven stress on the transportation network,[19] with congestion in one direction at the beginning of the workday and congestion in the other direction at the end of the workday – or in both directions simultaneously. In the base year, 2.37 million commuters crossed at least one RPA boundary on average each working day on their way to work, and then did it again on the way home.
- Alpha can change over time. For example, the trend towards remote work could shrink alpha towards 1 in many regions.
- Regions can be ranked by alpha.

Table 11 also includes this alpha measure for the RPA regions. The variation across RPAs in alpha is considerable, varying from 1.44 in the FRCOG RPA to .81 in the MAPC RPA. In every RPA except MAPC $\alpha > 1$, meaning that net out-commuting is positive – fewer people working in the region than workers living in the region. The MAPC RPA, on the other hand, is the hub of economic activity, jobs, and output in the state. This large region includes much of the Boston metropolitan area and draws in workers from surrounding regions and from outside of the state. Roughly half (49%) of resident Massachusetts workers live in this RPA, and nearly 60% (59%) of persons working in the state work in this RPA. As mentioned above, net in-commuting to the MAPC RPA is greater than 400,000 daily on average.

[19] An $\alpha = 1$ does not imply that there is *no* commuting; only that *net* commuting is zero. However, as α diverges from 1, commuting flows tend to be larger, resulting in more stress on transportation networks.

For the state as a whole, $\alpha = .97$, consistent with the view that Massachusetts, particularly the Boston metro area and the MAPC RPA, is the principal contributor to the New England gross domestic product.

Step 4: Transform resident labor force employment to payroll jobs by place of work

The end stage output requirements for the employment phase of the Socio Economic Projections is payroll jobs by 10 industry super sectors by place of work (RPA), the ES-202 employment concept. The previous step provides the number of jobs by detailed industry. To obtain the final payroll jobs estimates, the following transformations were made:

1. Jobs by place of work were converted to payroll jobs by place of work by dividing by the $1 + \gamma$ parameter. This factor nets out self-employment jobs (see Employment Concepts).
2. For 2020, the LODES-based payroll jobs by place of work estimates were calibrated to the 2020 ES-202 baseline jobs counts. A separate calibration factor is calculated for each industry super sector by RPA combination. Call these β factors.
3. For 2030, 2040, 2050, the LODES-based payroll jobs by place of work projections are factored by the β factors to provide the final payroll jobs estimates by place of work (RPA).

These β factors are in Table 12. The average β adjustment factor is 1.022, that is, the aggregate 2020 ES-202 job count exceeded the LODES estimate by 2.2%. The factor varies by RPA. For MAPC, the largest region, $\beta = 1.022$. The largest deviations of the β factors from 1 in magnitude were for the Nantucket Planning & Economic Development Commission (NPEDC), 17.8%; Old Colony Planning Council (OCPC), 7.0%; and Martha's Vineyard Commission (MVC), -13.6%. Several circumstances may have played a role in the size of these adjustments:

1. NPEDC and MVC are small regions – therefore subject to larger sampling error. They also have large seasonal variations in jobs.
2. The COVID recession in 2020 was an unusual year in terms of employment and may also have affected data collection and the accuracy of the data in 2020.
3. Differences in assigning industry codes. The ES-202 uses administrative records of state and federal agencies for coding industries, while the Census-based survey estimates rely on user-provided descriptions of their employer, sector, and type of work (U.S. Census Bureau, 2014; and Thompson, Kornbau, and Vesely, 2012). The reweighting in step 2 should mitigate this problem but does not do away with it.
4. The LODES data used were for 2015-19, which could be significantly different from the actual conditions in 2020, on which the ES-202 data are based.

The final baseline and projected payroll job counts by industry super sector and RPA are given in Table 1.

References

Clayton-Matthews, Alan. (2019). “Comparison of Least Squares Reweighting to Iterative Proportional Fitting”, memo, August 22, 2019.

Kevin S. Dubina, Janie-Lynn Kim, Emily Rolen, and Michael J. Rieley. (2020). "Projections overview and highlights, 2019–29," Monthly Labor Review, U.S. Bureau of Labor Statistics, September 2020, https://doi.org/10.21916/mlr.2020.21

Kevin S. Dubina, Lindsey Ice, Janie-Lynn Kim, and Michael J. Rieley. (2021). "Projections overview and highlights, 2020– 30," Monthly Labor Review, U.S. Bureau of Labor Statistics, October 2021, https://doi.org/10.21916/mlr.2021.20.

LODESTechDoc7.5.pdf (lehd.ces.census.gov/data/lodes/LODES7/LODESTechDoc7.5.pdf).

Thompson, Matthew, Micael Kornbau, and Julie Vesely. (2012). “Creating an Automated Industry and Occupation Coding Process for the American Community Survey.” In JSM Proceedings, Section on Statistical Learning and Data Mining. San Diego, CA: American Statistical Association. https://apps.bea.gov/fesac/meetings/2014-06-13_thompson_kornbau_vesely.pdf .

U.S. Census Bureau. (2014). American Community Survey Design and Methodology (January 2014), Version 2.0, January 30, 2014. https://www.census.gov/programs-surveys/acs/methodology/design-and-methodology.html .

U.S. Census Bureau. Public Use Microdata Sample (PUMS): Accuracy of the Data (2015-2019). https://www2.census.gov/programs-surveys/acs/tech_docs/pums/accuracy/2015_2019AccuracyPUMS.pdf

Williamson, P., Birkin, M., and Rees, P.H. (1998). The estimation of population microdata by using data from Small Area Statistics and Samples of Anonymized Records. Environment and Planning A, Vol. 30, No. 5, pp. 785-816.