FINAL REPORT

2013-2019 Massachusetts Residential Customer Profile Study

Massachusetts Program Administrators

Date: October 12, 2021

Table of contents

1 EXECUTIVE SUMMARY 1
1.1 Key findings 1
1.2 Recommendations 3

2 INTRODUCTION 4
2.1 Study purpose and objectives 4
2.2 Organization of report 4

3 METHODOLOGY 6
3.1 Data preparation 6
3.1.1 Data sources 6
3.1.2 Data modifications 7
3.1.3 Analysis variables 8
3.2 Improved renter identification 11
3.2.1 High-turnover flag 12
3.2.2 Owner name and use code flag 12
3.2.3 Approximate string-matching flag 12
3.2.4 College flag 12
3.2.5 Validation 12
3.3 Improved building type classification 13
3.4 Block group-level modeling 14
3.5 Individual modeling 15
3.5.1 Initial modeling 15
3.5.2 Data preparation 15
3.5.3 New modeling 15
3.6 Provisions 16
3.6.1 Data preparation 16
3.6.2 Renter confidence 16
3.6.3 Participation metrics 16
3.6.4 Model results 17

4 ANALYSIS AND RESULTS 18
4.1 Consumption, participation, and savings summary 18
4.1.1 Time-series summaries of locations and consumption 18
4.1.2 Location-level participation and savings 19
4.1.3 Program-level participation and savings 26
4.2 End-use analysis 29
4.2.1 Electric end uses 30
4.2.2 Gas end uses 32
4.3 Geographic market analysis 35
4.3.1 Electric geographic analysis 37
4.3.2 Gas geographic analysis 39
4.4 Renter Identification 41
4.5 Multifamily identification 43
4.6 Modeling 44

4.6.1 Block group-level models 44
4.6.2 Individual-level models 72

5 APPENDICES 80
5.1 Full correlation matrices 80
5.2 Input value definitions 84
5.3 Metrics definitions 84
5.4 RCPS analysis grains definitions 85

List of figures

Figure 4-1. Number of new and repeat participating locations between 2014 and 2019 by fuel. 23
Figure 4-2. Trend for electric population savings across Residential New Homes, Residential Existing Buildings, and Income Eligible Existing Buildings from 2013-2019 28
Figure 4-3. Comparison between savings and incentives contribution for electric measures 30
Figure 4-4. Comparison between savings and incentives contribution for electric delivery pathways 31
Figure 4-5. Electric savings by year and channel 32
Figure 4-6. Comparison between savings and incentives contribution for gas measures 33
Figure 4-7. Comparison between savings and incentives contribution for gas delivery pathways 34
Figure 4-8. Gas savings by year and channel 35
Figure 4-9. Population size represented by the number of occupied housing units by town 36
Figure 4-10. Map of incentive dollars per kWh consumption by town, cumulative 2013-2019 37
Figure 4-11. Map of incentive dollars per kWh consumption by town, 2018 38
Figure 4-12. Map of incentive dollars per kWh consumption by town, 2019 38
Figure 4-13. Map of incentive dollars per therm consumption by town, cumulative 2013-2019 39
Figure 4-14. Map of incentive dollars per therm consumption by town, 2018 40
Figure 4-15. Map of incentive dollars per therm consumption by town, 2019 40
Figure 4-16. Block group renter percent; DNV method vs. ACS 42
Figure 4-17. Correlation of electric location-level participation with ACS/US Census variables 49
Figure 4-18. R^2 progression of best-fit model – electric location participation 50
Figure 4-19. Correlation of electric consumption-weighted participation with demographic variables 54
Figure 4-20. Correlation of electric population savings rate with demographic variables 57
Figure 4-21. R^2 progression of best-fit model – electric population savings rate 58
Figure 4-22. Correlation of gas location-level participation with ACS/US Census variables 62
Figure 4-23. R^2 progression of best-fit model – gas location participation 63
Figure 4-24. Correlation of gas consumption-weighted participation with demographic variables 66
Figure 4-25. Correlation of gas population savings rate with demographic variables 69
Figure 4-26. R^2 progression of best-fit model – gas population savings rate 70
Figure 4-27. Building size * Ownership interaction – electric location savings rate 77
Figure 4-28. Building size*Ownership interaction – gas location savings rate 79
Figure 5-1. Correlation of participation metrics with demographic variables – electric 82
Figure 5-2. Correlation of participation metrics with demographic variables – gas 83

List of tables

Table 1-1. 2013-2019 RCPS key findings and observations 1
Table 3-1. RNCPS renter flag inaccuracies 11
Table 4-1. Number of unique physical locations by year, 2013-2019 18
Table 4-2. Annual consumption electric (MWh) and gas (Dth), 2013-2019 19
Table 4-3. Statewide savings for upstream, behavioral, and all other programs for 2013-2019 19
Table 4-4. Participant savings (MWh, Dth) with behavioral and with upstream by year, 2013-2019 20
Table 4-5. Participant savings (MWh, Dth) without behavioral and with upstream by year, 2013-2019 20
Table 4-6. Participant savings (MWh, Dth) without behavioral and without upstream by year, 2013-2019 21
Table 4-7. Participation rates and population savings achieved by year, without behavioral and with upstream savings, 2013-2019 21
Table 4-8. New and repeat participating locations, without behavioral and without upstream 22

Table 4-9. Time-series participation rates for unique electric and/or gas PA-served locations, without upstream or behavioral. 24
Table 4-10. Time-series participation rates for unique electric PA-served locations, without upstream or behavioral. 24
Table 4-11. Time-series participation rates for unique gas PA-served locations, without upstream or behavioral. 25
Table 4-12. Participation rates by PA without behavioral and without upstream, cumulative from 2013 25
Table 4-13. Participation rates and population savings achieved by PA, without behavioral and with upstream savings, 2019 26
Table 4-14. Participation rates and population savings achieved by PA, without behavioral and with upstream savings, 2018 26
Table 4-15. Participant locations served by program, without behavioral and without upstream, 2013-2019 27
Table 4-16. Population savings by program, without behavioral and with upstream, 2013-2019 27
Table 4-17. Population savings by program, without behavioral and without upstream, 2013-2019 27
Table 4-18. Population savings by PA and program without behavioral and with upstream, 2019 28
Table 4-19. Population savings by PA and program without behavioral and with upstream, 2018 29
Table 4-20. Examples of measures included in broad end use categories. 29
Table 4-21. Statewide renter proportions 41
Table 4-22. Electric PA renter proportions 41
Table 4-23. Gas PA renter proportions 41
Table 4-24. Comparison of housing unit classifications between DNV's algorithm and the ACS estimates 43
Table 4-25. Comparison of renter ratios by building size category between DNV and the ACS 44
Table 4-26. Key takeaways from the block group-level modeling - electric 45
Table 4-27. Key takeaways from the block group-level modeling - gas 46
Table 4-28. Best-fit model – electric location participation 50
Table 4-29. Model A – Term sheet and multifamily – Electric location participation 52
Table 4-30. Model B – Multifamily without Renter – Electric location participation 53
Table 4-31. Model C – Effect of all descriptive variables – Electric location participation 53
Table 4-32. Best-fit model – electric consumption-weighted location participation 55
Table 4-33. Model A – Term sheet and multifamily – Electric consumption-weighted location participation 55
Table 4-34. Model B – Multifamily without Renter – Electric consumption-weighted location participation 55
Table 4-35. Model C – Effect of all descriptive variables – Electric consumption-weighted location participation 56
Table 4-36. Best-fit model – electric population savings rate 58
Table 4-37. Model G – Term sheet, low income, and multifamily – Electric population savings rate 59
Table 4-38. Model H – Term sheet, low income, and multifamily – Electric population savings rate 60
Table 4-39. Model C – Effect of all descriptive variables – Electric population savings rate 60
Table 4-40. Best-fit model – gas location participation 63
Table 4-41. Model A – Term sheet and multifamily – Gas location participation 64
Table 4-42. Model B – Multifamily without Renter – Gas location participation 64
Table 4-43. Model C – Effect of all descriptive variables – Gas location participation 65
Table 4-44. Best-fit model–gas consumption-weighted location participation 66
Table 4-45. Model A – Term sheet and multifamily – gas consumption-weighted location participation 67
Table 4-46. Model B – Multifamily without Renter – gas consumption-weighted location participation 67
Table 4-47. Model C – Effect of all descriptive variables – gas consumption-weighted location participation 68
Table 4-48. Best-fit model – gas population savings rate 70
Table 4-49. Model G – Term sheet, low income, and multifamily – gas population savings rate 71
Table 4-50. Model H – Term sheet, low income, and multifamily – gas population savings rate 72
Table 4-51. Model C – Effect of all descriptive variables – gas population savings rate 72
Table 4-52. Individual-level modeling key takeaways – electric 73
Table 4-53. Individual-level modeling key takeaways – gas 74
Table 4-54. Location-level probability of participation - electric 75
Table 4-55. Frequency of significant prediction of location savings rates – electric 76
Table 4-56. Mean location savings rate by category - electric 76
Table 4-57. Location-level probability of participation – gas 77
Table 4-58. Frequency of significant prediction of location savings rate - gas 78
Table 4-59. Mean location savings rate by category - gas 79
Table 5-1. Full correlation matrix - electric 80
Table 5-2. Full correlation matrix - gas 81

1 EXECUTIVE SUMMARY

The 2013-2019 Residential Customer Profile Study (RCPS) analyzes and reports on the Massachusetts Program Administrators' (PAs') energy efficiency program tracking and residential and income-eligible customer consumption data enhanced with third-party demographic data to offer diverse views of participation, savings, and geographic dynamics within the PAs' residential customer population.

1.1 Key findings

Table 1-1. 2013-2019 RCPS key findings and observations

Key finding No.	Report section	Key finding
1	**Time-series summaries of locations and consumption**	Savings gains from electric programs appear to be outpacing population growth, with a 7% growth in locations served between 2013-2017 accompanied by only a 1.6% increase in electric consumption.
2	**Location-level participation and savings**	The PAs' programs continue to see increasing savings, as savings across all programs and both fuels have followed an upward trend since 2013. Annual savings for all electric programs have increased by 415,492 MWh (a 67% increase) between 2013 and 2019. Annual savings for all gas programs have increased by 474,069 Dth from 2013 through 2019 (a 22% increase).
3	**Location-level participation and savings**	There is a positive trend between 2013 and 2019 in the number of repeat participants for both gas (a 220% increase) and electric (a 247% increase) programs. This finding suggests that the PAs are encouraging customers to return to their program offerings at increasing rates.
4	**Location-level participation and savings**	From 2013-2019, electric location-level participation rates averaged around 7% annually. Gas location-level participation averaged around 5%. The 7-year average annual participation rate across both electric and gas is approximately 9%, and over this 7-year period, 40% of locations participated (upstream lighting not included).
5	**Renter Identification**	A key conclusion of the RNCPS was that renter status is the most important and practical marker for the PAs to identify who is likely to be underserved. Utility records do not indicate if a customer is an owner or renter, so we have developed a renter identification approach that has a very strong correlation (0.75) with the ACS data.

Key finding No.	Report section	Key finding
6	**Block group-level modeling (electric, gas)**	The term sheet[1] requires the PAs to assess how well they are providing energy efficiency services to renters, moderate-income customers, and customers with limited English proficiency. One of the outcomes of the previous report was that renter status was the most predictive characteristic of participation. Renter-occupied housing is the strongest single demographic predictor of electric location participation, gas location participation, and gas population savings rates. Single-family housing is the strongest single demographic predictor of electric savings rate.
7	**Block group-level modeling (electric, gas)**	The best combination of variables predicts approximately 23% of the variance of electric participation (24% of gas). In other words, approximately three-fourths of the participation is driven by factors other than those considered in this study. This implies that interventions other than simply targeting demographics might be effective at improving participation rates.
8	**Block group-level modeling (electric, gas)**	Renting and lower-than-average income are the drivers of participation in areas with high concentrations of limited English proficiency households. By itself, the correlation between limited English proficiency and electric participation is moderately strong at -0.23 (-0.26 for gas). When renting and income are added to a regression model with limited English proficiency, the latter effect became non-significantly greater than zero.
9	**Block group-level modeling (electric, gas)**	Even when controlling for the overlaps among demographic variables, areas with high concentraitons of households with limited English proficiency and/or moderate incomes tend to have lower savings rates. When these customers participate, they appear to be achieving lower savings than those without these demographic characteristics.
10	**Block group-level modeling Individual-level modeling (electric, gas)**	The block group-level modeling results are consistent with the available individual-level modeling results (building size, ownership, and building age). This adds justification for the PAs to act on the block group-level results, with caution.
11	**Individual-level modeling (electric)**	Building size was the best predictor of electric participation and depth of savings at an individual location level. Single family buildings are the most likely to participate in electric programs (33%) and 5–9-unit buildings are the least likely to participate in electric programs (17%). When they do participate, single-family homes have the greatest depth of savings (savings

[1] http://ma-eeac.org/wordpress/wp-content/uploads/Term-Sheet-10-19-18-Final.pdf

Key finding No.	Report section	Key finding
		rate =18.7) and 5–9-unit buildings have the least depth of savings (savings rate=12.9), followed closely by triplex/fourplex buildings (savings rate=13.1).
12	**Individual-level modeling (gas)**	For gas participation rates, building size is a strong predictor of participation rates. Single-family homes are the most likely to participate (24%) and 5–9-unit buildings are the least likely (8%). Ownership also predicted participation with owners more likely to participate (23%) than renters (17%). For the depth of gas savings of participating buildings, ownership is the most consistent predictor, with renters achieving greater savings rates (14.0) than owners (11.2). This effect appears to be driven by an interaction between building size and ownership. There is little difference in gas savings rates for participating buildings that are single family, duplex, or triplex/fourplex. However, for 5+ unit buildings, participating renters have a greater depth of gas savings than owners.

1.2 Recommendations

Recommendation 1: Most lighting will become ineligible for program funding in the next few years. When this occurs, the PAs will find it increasingly difficult to achieve overall program savings goals and pass cost-effectiveness tests. The PAs should plan for another measure or program approach to replace the savings from lighting. Lighting accounts for most electric savings, and it has one of the highest benefit-cost ratios of all electric measures.

Recommendation 2: The PAs should continue focusing on renters for targeted program marketing. For both gas and electric programs, the PAs should specifically increase targeting 5-9-unit buildings. These buildings had the lowest participation rates and savings rates across the different categories considered.

Recommendation 3: The PAs should investigate ways to increase depth of savings for moderate-income customers and customers with limited English proficiency. Areas with high concentrations of these characteristics had a lower depth of savings.

Recommendation 4: The PAs should continue using the readily available block group-level demographic results for program targeting and other implementation purposes. With the ability to produce individual-level variables for building size, ownership, and building age based only on PA and tax assessor data, evaluators should continue to consider both the block group- and individual-level analyses for investigation, recommendations, and confirmation of the block group-level analyses.

2 INTRODUCTION

2.1 Study purpose and objectives

The purpose of the 2013-2019 Residential Customer Profile Study (2013-19 RCPS) is to characterize the Massachusetts residential energy efficiency market by analyzing the usage, savings, and program participation data of all residential gas and electric customers served by the Massachusetts Program Administrators (PAs). The primary objective of the study is to identify historical trends in participation metrics, their potential drivers, and their future implications for PA customer populations. Other objectives are to integrate this study with the Residential Customer Profile Study Dashboard and identify possible areas of further research to inform the future direction of the PAs' programs.

This study integrates and expands on several recent profile customer reports DNV has completed for the Massachusetts PAs:

- **2013-2017 Residential Customer Profile Study (13-17 RCPS)** – This study provides summary statistics of the PA customer and participant base from 2013-2017. It includes several analyses exploring the relationships between specific demographics and participation.
  The 2013-19 RCPS covers the same summary statistics as the 2013-17 RCPS, expanding them to include 2018 and 2019. In addition, this report repeats the demographic analyses to continue providing insights into the participation patterns of various demographic groups.
- **2013-2017 Residential Nonparticipant Customer Profile Study (RNCPS)** – This study explores the relationships between participation and several specific demographics of interest (moderate income, renters, and people with limited English fluency) using different methods from the 13-17 RCPS. It includes bivariate and multivariate analyses at the block group level. It also includes bivariate and multivariate analyses at the individual account level for Eversource customers.
  The 2013-19 RCPS expands on the RNCPS by adding data from 2018 and 2019 to the analyses, and integrates the bivariate, block group-level analyses into the same report as the rest of the residential customer profiling. In the 2013-2019 report, we also explore improvements to assigning renter status to individual accounts based on information available from all PAs and conducting individual-level, multivariate analyses involving these data.
- **RCPS Dashboard** – This project provides many of the same statistical summaries reported in the 2013-17 RCPS, adding in 2018 data, and makes them all available on an interactive dashboard.
  Under the 2013-19 RCPS, DNV updated the RCPS Dashboard with 2019 data, and greatly expanded the socioeconomic sections to include third-party data sources such as the American Community Survey (ACS), Housing and Urban Development (HUD), and Massachusetts Level 3 tax assessment data.

2.2 Organization of report

The remainder of this report is organized as follows:

Section 3: Methodology

Section 4: Analysis and results

- Section 4.1: Consumption, participation, and savings summary
- Section 4.2: End-use analysis
- Section 4.3: Geographic market analysis
- Section 4.4: Renter identification

Section 4.5: Multifamily identification

Section 4.6: Modeling

Section 5: Appendices

Section 5.1: Full correlation matrices

Section 5.2: Input value definitions

Section 5.3: Metrics definitions

3 METHODOLOGY

This report provides analyses that complement and expand on those in the RCPS Dashboard using the PAs' most recent program tracking and billing data for 2018 and 2019 as well as historic data from 2013-2017.

3.1 Data preparation

The RCPS uses the PAs' 2013-2019 billing and tracking data after it has been through the extract, transform, and load (ETL) process under the 2019 Residential Data Intake project and signed off by the PAs via a Summary of Data Completeness (SDC) report to each PA.

3.1.1 Data sources

3.1.1.1 PA billing and tracking data

After the PAs' customer billing and energy efficiency program tracking data has been through the ETL process, it is incorporated into the MA Data Warehouse, where it becomes the foundation of the RCPS. The MA Data Warehouse is a time-series database with residential data beginning in 2013. It is designed to support multiple research and evaluation projects, to allow research teams to design and select samples, and to efficiently estimate population results. It also allows for evaluation contractor teams to receive approved and cross-checked data to support ongoing evaluation work and to respond to time-sensitive ad hoc requests from the PAs and EEAC Consultants. The current time window for which PA data is available is 2013 through 2019; DNV has obtained and processed all PA billing and tracking data for that period.

3.1.1.2 Third-party data

DNV supplements the PA billing and participant tracking data aggregated in the MA Data Warehouse with several third-party datasets containing a series of demographic, financial, and building characteristic attributes at either the detailed location level (Tax data and e911 data) or block group level (ACS data). Key datasets included in the MA Data Warehouse are the Emergency 911 location data, the MA Level 3 tax assessment data, and the ACS data. These are described below.

The **Emergency 911 geographic data** is a detailed list linking address information (for houses, buildings, apartment blocks, etc.) to corresponding latitudes and longitudes. This source allows us to place nearly every PA customer at a precise geographic location, so we know where, geographically, program activity is occurring. This includes placing facility-level tracking data without account numbers, and C&I data for multifamily locations and housing authorities.

The **MA Level 3 tax assessment data** comes from each town's tax assessor. It contains information about sales, building type, building age, and ownership, and classifies every building into one of several hundred building use codes (like "apartments with 4-8 units" and "residential condominium"). To make this data practical, we simplify these detailed codes into 9 main building types:

1. Single Family
2. Duplex/Triplex/Condo
3. Small Multifamily, 4-8 Units
4. Large Multifamily, 8+ Units
5. Mobile/Other Residential
6. Housing Authorities/Subsidized Housing
7. Mixed Use – Primarily Residential
8. Other Non-Residential
9. Unknown

With these 9 main types, we can show each PA the makeup of its customer and participant building population.

ACS data is collected by the US Census Bureau. It includes numerous details such as languages spoken, household income, travel times, building age, and building ownership. Notably, these details are not publicly available at the building level, but at the level of the block group—a unique group of roughly 200 to 1,000 households covering a contiguous area. The 2013-2019 RCPS uses the 5-year rolling average ACS data covering 2015-2019.

3.1.2 Data modifications

The data preparation followed the convention adopted last year of using the physical location as the RCPS's analysis grain. Physical location best aligns across the population of PA tracking and billing systems, is generally a more static analysis grain than customers or accounts over time, and facilitated clean aggregation of data across fuels, PAs, and with third-party data.

In this study, DNV reports on the unique physical locations represented in the PAs' data. DNV identified unique locations using a series of logic models previously developed for the PAs, complemented by the MA Level 3 tax assessment data. We accomplished this by:

1. Standardizing the PA-provided physical location data at the address level (removing sub-unit information such as apartment number, floor, etc.) and assigning a unique physical address ID to each physical location
2. Geocoding the physical address using the Emergency 911 geographic data, TIGER, and other geocoder processes to assign a unique geographic coordinate ID. This helped resolve instances of single buildings having multiple addresses associated with them, such as a duplex home representing addresses 2 and 4 Main St.) Only addresses that successfully geocode are counted (this makes up roughly 95% of the population of addresses).
3. Deploying the previously developed logic models to update and match a tax parcel ID with any linked property assessment IDs, when possible and appropriate

Data additions: The PA implementation staff indicated that when multifamily properties, particularly low-income properties, receive energy efficiency measures, the entire building is treated at the same time and all sub-units are impacted. Additionally, low-income multifamily properties that may have some consumption on a C&I rate—for example, a centralized heating system—participate through the PAs' Low-Income program, and an accurate reflection of participation and savings would need to include the C&I consumption element for these low-income multifamily properties (i.e., housing authorities and public housing). Using building type information from tax data and a string search on customer names (e.g., names that included various permutations of "HOUSING AUTHORITY"), we identified housing authorities and subsidized multifamily housing in the C&I billing data. These records were added to the residential billing data to account for the fact that buildings with C&I accounts housing low-income tenants participate through the PAs' low-income initiatives.

Data removals: Consistent with the 2013-2017 RCPS, DNV excludes behavioral and upstream data from select tables and charts; these exclusions are noted throughout the report. The unique factors behind instances where these programs are excluded from an analysis output are:

- **Behavioral data:** The behavioral program is a passive engagement offering. Customers are enrolled as either a treatment or a control, and "treatment" customers receive a mailing attempting to modify their behavior regarding energy use. Including the accounts in the behavioral program would cause participation rates for PAs that have a behavioral program to appear artificially high due to the passive participation; for some of the more detailed geographic analyses, block groups would appear to have 100% participation. Including the behavioral savings would also make it difficult to have apples-to-apples comparisons between the PAs that do and do not offer behavioral programs.
- **Upstream data:** This data is reported at the point-of-sale (store) level and does not include any purchasing customer information; this means it is not possible to identify individual accounts, customers, or locations that participated in the program. The implication is that the upstream program does not impact any calculated participation rates in the RCPS.

Savings are allocated geographically using a model created by the NMR team that is based on the American Community Survey (ACS) household numbers and characteristics. This does allow for the spatial representation of where those savings occur, but because those savings are perfectly correlated with the ACS variables used to allocate them, the upstream savings cannot be included in the demographic analyses and modeling.

- **Outlier block groups:** For the modeling tasks only, DNV conducted several additional tests to remove block groups that represent edge cases and outliers. First, any block group with fewer than 25 accounts in the consumption data was removed. These are assumed to be on the edge of PA service territories where most of the block group is served by a municipal or other non-PA provider. Because of the small amount of PA-based consumption in these block groups relative to the full size of the block group, metrics such as participation rates and savings rates tend to be unstable. DNV conducted additional data cleansing for the models based on the independent variables, which are described in the next section.

3.1.3 Analysis variables

The three main variables used in this report are defined as follows:

A **location** is defined as a unique combination (concatenation) of the PA-provided address, zip code, and town/city name. Locations tend to be more static than accounts, which change when occupants move in or out of a specific residence.

Participation is represented in three different ways in this report. All three participation variables are calculated for:

- A specific geographic area, typically an entire PA service territory or a specific block group
- During a particular period, typically a specific year or all the years in the analysis window (2013-2019)

Location-level participation rate is calculated as the number of participating locations divided by the total number of locations. Location participation rate represents the breadth of program coverage.

The location-level participation rates used in the modeling task are calculated at the individual block group level for the entire analysis period of 2013-2019:

$$LocationParticipation_{blockgroup} = \frac{\sum_{2013}^{2019} Participating\ Locations_{blockgroup}}{\sum_{2013}^{2019} All\ Locations_{blockgroup}}$$

Population savings rate is defined as the sum of savings from participating locations divided by the sum of consumption for all locations. The population savings rate represents the depth of program coverage.

The population savings rates used in the modeling task are the sum of 2013-2019 savings divided by the average of 2013-2019 consumption taken over years when consumption was greater than 0.[2] Both are calculated at the block group level.

$$PopulationSavingsRate_{blockgroup} = \frac{\sum_{2013}^{2019} Savings_{blockgroup}}{\frac{1}{n} * \sum_{2013}^{2019} Consumption_{blockgroup}}$$

Where n is equal to the number of years where $Consumption_{blockgroup} > 0$

Consumption-weighted location participation is defined as the sum of consumption of participating locations divided by the sum of consumption of all locations. The consumption-weighted participation rate is an alternative way to represent the breadth of participation.

[2] Using the average yearly consumption reduces the influence of year-to-year fluctuations in consumption. DNV conducted sensitivity analyses that showed that correlations and models using the sum of 2019 savings in the denominator were very consistent with those that use the 7-year average.

For the modeling task, consumption-weighted participation was calculated as (the sum of the 2019 savings for any location that participated at least once between 2013 and 2019) divided by (the sum of the 2019 consumption of all locations).

$$ConsumptionWeightedParticipation_{blockgroup} = \frac{\sum_{2019}^{2019} Consumption\ of\ Participating\ Locations_{blockgroup}}{\sum_{2019}^{2019} All\ Locations_{blockgroup}}$$

Additional data trimming used in the modeling tasks includes the following:

- For analyses using location-level participation rate and consumption-weighted participation rate, any block group with a location-level participation rate or consumption-weighted participation rate greater than 100% was removed from the analysis. This could happen for multiple reasons, typically related to incomplete or incorrect address information (for example, as an effect of incorrect block group assignments—accounts placed outside their real location by the geocoder, or the number of accounts in the denominator being too small when multiple locations are treated as one).
- For analyses using population savings rate, any block groups with a population savings rate less than 0% or greater than 100% were removed from the analyses. Negative savings can occur when PAs enter a tracking record with negative savings to make a correction or true-up savings. Population savings rates greater than 100% are theoretically possible because the numerator represents seven years of savings and the denominator only one year. However, it is extremely rare for this to occur. The most likely cause of this pattern is substantial new construction occurring in a block group.

The remainder of the variables defined in this section were used as explanatory variables in the modeling tasks.

Moderate income: In defining moderate income, we note that the ACS and HUD do not provide data that directly match the state definition of a household income between 56% and 85% of the statewide median. Thus, for any approach considered, there will be households excluded which fit the state's definition and households included that do not fit it, which introduces uncertainty. DNV defines a moderate-income variable as ACS variables (B19001e9 + B19001e10 + B19001e11) / B19001e01, which can be read as percentage of households with incomes between $40,000 and $59,999. This is equivalent to 53.3%-80.0% of state median income. This definition matches the one used in the RNCPS. DNV considered several other ways of defining this variable, and all definitions had similar behavior in the correlation and modeling results. Thus, in the interest of consistency, we decided to retain the same definition used previously.

Mass Save defines households with an income between 56% and 85% of statewide median income as moderate-income households. According to the latest ACS data,[3] statewide median income in Massachusetts for the 2015-2019 period is approximately $81,000. Thus, moderate-income is approximately $45,000 to $65,000. However, the Mass Save definition also considers a second variable: the number of household members. Including this variable allows for fine-tuning low- and moderate-income definitions to consider the higher expenses incurred by larger households, which likely include (non-working) children. Thus, the low- and moderate-income thresholds for single-person households are lower than those for households with 5 or more family members. Thus, the true range of moderate-income for Mass Save[4] is from just over $40,000 for single-member households to just over $173,000 for households with 17 members.

The ACS provides household income data in bins with thresholds that do not exactly match these end points. The bins in the ACS data closest to the moderate-income range are:

- $40,000 to $44,999
- $45,000 to $49,999

[3] https://www.census.gov/quickfacts/fact/table/MA/INC110219

[4] https://www.masssave.com/saving/income-based-offers

- $50,000 to $59,999
- $60,000 to $74,999
- $75,000 to $99,999
- $100,000 to $124,999
- $125,000 to $149,999
- $150,000 to $199,999

Using the ACS variable to define moderate-income requires choosing which of these bins to include. The closest endpoints to statewide median income (independent of household size) in the ACS data are $45,000 to $60,000. Considering the possibility of single-member household thresholds going as low as $40,000 could justify the inclusion of the previous bin. Including the $60,000 to $74,999 bin could be justified to capture the top of the moderate-income range. DNV tested the behavior of a moderate-income definition using $45,000 to $74,999 and it acted very similarly to the $40,000 to $59,999 definition that we ultimately used.

The Department of Housing and Urban Development (HUD) maintains a low- and moderate-income database that considers both ACS income and household size to report the number of low- and moderate-income individuals in a block group.[5] The HUD data defines low income as less than 50% of the area median income, and moderate-income as 50% to 80% of the area median income. Thus, using this dataset allows for a more precise match to the two-variable moderate-income definition used by Mass Save. Limitations of this dataset are:

- It uses area median income as a basis rather than state median income
- The most recently available data is based on the 2011-2015 ACS 5-year rolling average
- It reports at the grain of individuals, not households.

DNV tested the behavior of a moderate-income definition based on the HUD data, and it acted very similarly to the $40,000 to $59,999 definition that we ultimately used.

Low income: The definition for low income is households earning less than $40,000 based on the ACS data. Specifically, this variable is calculated by summing the ACS variables B19001e02 to B19001e08 and dividing the sum by B19001e01. This reflects the households with less than 53.3% average 2013-2019 statewide median income (and less than 46.6% of 2019 statewide median income).

The previous discussion about different ways to define moderate-income apply to low income as well. As with moderate income, DNV tested the behavior of a low-income variable based on the HUD data, and it behaved much like the ACS-only variable we ultimately used.

Average or higher income: DNV defined average or higher income as the proportion of households with greater than Moderate income (based on ACS variables and calculated as the [sum of B19001e13 to B19001e17] divided by B19001e01).

Renter-occupied housing: DNV calculated this variable as the proportion of renter-occupied households in the block group (calculated as B25003e3 / B25003e1).

Multifamily housing: DNV calculated this variable as the proportion of occupied households in the block group living in buildings with 5 or more units (based on ACS variables in table B25024 and calculated as (B25024_e06 + B25024_e07 + B25024_e08 + B25024_e09) / B25024_e01).

[5] https://www.hudexchange.info/programs/acs-low-mod-summary-data/

Limited English-speaking households: DNV calculated this variable as the proportion of households in the block group reporting as speaking limited English at home (based on ACS variables in table C16002 and calculated as (C16002_e03 + C16002_e06 + C16002_e09 + C16002_e012) / C16002_e01).

Construction year pre-1950: DNV calculated this variable as the proportion of housing units built before 1950 (based on ACS variables in table B25034 and calculated as (B25034_e10 + B25034_e11) / B25034_e01).

Urban PA household ratio: The most recent data for block-level urban or rural classification comes from the 2010 U.S. Census. The Census reports urban and rural classifications at the level of the block, rather than the block group. Thus, DNV had to aggregate the results up to the block group level. This required weighting each block according to the proportion of the population represented in the entire block group. To do this, DNV used the number of PA accounts that are served in the block. We then calculated the urban ratio for each block group as the number of accounts in urban blocks within this block group divided by the number of all accounts served in that block group.

3.2 Improved renter identification

One of the key conclusions of the RNCPS was that the renter variable was the most important and practical marker for the PAs to identify who is likely to be underserved. The RNCPS used a simple method with only PA data to assign probable renter status at the location level. For any non-condo location that had 3 or more accounts over the 5-year evaluation window, DNV assigned renter status to all accounts at that location. This had a 90% match rate with Experian data, which uses a more robust method that combines statistical interpolation with primary research. However, because Experian data was only available for Eversource, the RNCPS analyses had to rely on our renter flag for everyone else.

The renter assignment method used in the RNCPS over- and under-counted renters in several situations. DNV used the tax assessment data to adjust the accounts in many such situations to reduce the number of inaccuracies (Table 3-1).

Table 3-1. RNCPS renter flag inaccuracies

Situation	Inaccuracy
Renter in same unit >3 years (including new construction)	Undercounted
Condos that are rented out	Undercounted
Multifamily condos without "condo" in account name	Overcounted
Account owner changes due to death, divorce	Overcounted
Spouses change name account is held under	Overcounted
Property sold many times during evaluation window	Overcounted

For the 2013-19 RCPS, DNV attempted to address some of these inaccuracies by making improvements to the renter identification process used in the RNCPS. We developed 4 flags that use data from the PAs, the tax assessment, and public data on university locations in Massachusetts. The flags are:

1. The high-turnover flag
2. The owner's name and use code flag
3. The approximate string-matching flag
4. The college flag.

A location or individual flagged by any of these 4 algorithms is a likely renter. Following is a description of the logic behind each algorithm.

3.2.1 High-turnover flag

The high-turnover flag was the only renter flag used in the RNCPS. It is based only on information available from the PAs. For this approach, we identified a high turnover rate (i.e., a higher number of accounts associated with a location from 2015-2019) as a proxy for renter status. We assumed that residents in rented properties move in and out more frequently than residents of non-rented properties. We set a threshold for the number of account changes to represent this tendency. However, an account change does not necessarily indicate a renter turnover or even a sale or change of ownership, and the threshold approach underrepresents long-tenured renters.

3.2.2 Owner name and use code flag

For this flag, we leveraged data from tax assessment records to identify likely renters. We performed partial string searches for terms indicating that a residential property is rented, including “COLLEGE,” “APARTMENT,” and “TENANT,” among others. We also leveraged building use codes from the tax assessment data to identify rental locations, including use codes for apartments or housing authorities. We also assumed that condos are owned and flagged these as non-renter locations by searching for permutations of the text string “CONDO” in the property owner's name and by using tax assessment building types corresponding to condos. This means that rented condos are underrepresented by this approach.

3.2.3 Approximate string-matching flag

This approach assumes that if the individual listed as the property owner is different than the individual in the PA’s billing records, the location is likely to be rented. However, tremendous variability in customer/owner name formats between the billing and tax data makes direct text string comparisons inaccurate. For example, “John Smith” and “John H. Smith” are likely the same person but would be flagged as distinct individuals using direct text string comparison. To overcome this limitation, we employed a “fuzzy matching” algorithm that measures the dissimilarity between two strings and assigns an edit-distance score based on the number of changes required to transform the comparator string into the comparison string.[6] Because this edit-distance score only indicates the relative similarity between two strings, DNV had to holistically choose an appropriate cut-off score for identifying matching customer/owner names calibrated by extensive spot-checking. Our approach here was conservative so that only names that differed by a large margin were flagged as renters. Additionally, if the customer’s name or owner’s name was missing from the tax data then they were identified as not a renter because this algorithm could not be applied to that location. This approach thus tends to under-count potential renter locations.

3.2.4 College flag

Locations near colleges and universities which house renters have a high likelihood of housing student renters, who are most likely to have rental agreements timed to the school year. We used this assumption to identify potential student renters by sub-setting the high-turnover flag to only locations within one mile of colleges and university locations based on public data. We then identified locations with account turnover in May, June, August, or September and flagged these as likely student renters. The DNV team used Massachusetts college location latitude and longitude coordinates to calculate the distances to the housing location latitude and longitude coordinates in the e911 data. These calculations were done separately for gas and electric.

3.2.5 Validation

Location-level information about renters in Massachusetts is very limited, so beyond spot-checking individual locations, we are limited to performing validation comparisons with other data sources at a coarser grain. The ACS provides block-group-level counts of renter-occupied housing units, so we compared our results at the block group level to the ACS counts, and used these results to guide modifications to the algorithms or to inform data investigations to identify and resolve anomalies. This approach is limited by the fact that the ACS includes information for municipally served territories for which we have no

[6] The specific function used is the “COMPGED” routine included in the SAS analytics software package.

PA data (approximately 15% of all housing units are in municipal territories), and where we do have PA data, we are limited by how successfully addresses can be geocoded to US census block groups (7% of locations are not geocoded to a block group).

3.3 Improved building type classification

During report development, stakeholders raised the issue that using tax assessment data directly to determine building type can be problematic because the tax assessment building size classifications are arbitrarily determined by each town and do not necessarily match US Census classifications. To overcome this limitation, DNV developed an approach to a "best-guess" estimate of building type classification based on all known information about a parcel. This approach leverages tax assessment data where it can be used and combines it with PA data to arrive at building-size classifications consistent with those used by the ACS. The classifications are:

- 1 unit
- 2 units
- 3-4 units
- 5-9 units
- 10-19 units
- 20+ units

This analysis used only electric location data because the higher percentage of master-metered locations in the gas data significantly undercount the number of units in large multifamily buildings. We also observed that while the tax assessment data often contains estimates of the number of units in a building along with a building type designation (e.g., "single family" or "4+ unit apartment building"), the number of units is often incongruous with the building type assignment (a two-family home showing 8 housing units, for example), making it necessary to implement additional logical steps to use this data in our estimates.

The algorithm has two stages: the first is a broad-brush assignment of the number of units in a building based on the number of unique secondary addresses identified at its unique primary street address based on the cleaned and geocoded PA billing data. The second leverages tax assessment data to refine the prior stage's estimates based on the following logic:

- If an address has a single secondary address, a single electric billing account, and is designated as a single-family home in the tax assessment data, it is estimated to be a single unit.
- If an address has 1-2 secondary addresses or 1-2 billing accounts in the electric billing data and is designated as a two-family home in the tax assessment data, it is estimated to be two units. This helps to refine the estimates of two-family homes with a single meter.
- If an address has 1-3 secondary addresses or 1-3 billing accounts in the electric billing data and is designated as a three-family home in the tax assessment data, it is estimated to be three units. This similarly helps to refine estimates of three-family homes with a single meter.
- If an address is designated as 4-8 units in the tax assessment data and the number of units in the tax assessment data is between 4 and 8 inclusive, the number of units in the tax assessment data is used as the estimate. Otherwise, if the number of units in the tax assessment data is greater than 8, the estimate is the mean of the number of units from the tax assessment data and the number of secondary addresses at the location from the PA data, or the estimate is the number of secondary addresses at the location. This helps to refine estimates where there are inaccurate counts of secondary addresses.
- If a location has a single billing account but 20 or more units in the tax data, the estimate is the number of units in the tax data. This helps to refine the estimate for master-metered large multifamily buildings.

There are substantial limitations to these approaches, some of which are noted above, so this analysis should be taken only as a "best estimate" of building type category based on the available information for a location. We are limited where we cannot append tax assessment information to the PA billing data using only the location information in the PA data (21% of locations in the PA data do not have tax data appended), and we are limited by the quality of the address information provided by the PAs. Inconsistency in reported secondary addresses, for instance, will introduce error into the count of units based on those addresses; a secondary address might appear variably within the same year's billing data as "Apt B" and "Apt 2," for instance, which will be counted as two distinct units. It is not possible to quantify the extent to which this scenario occurs, so it is a source of uncertainty for this method.

3.4 Block group-level modeling

Before modeling, DNV standardized all variables across the block groups, including the participation rates and the demographic variables. The standardization process converts the raw metric into a z-score, which centers the distribution's mean at zero and converts units into standard deviations from that mean. All these transformations are linear, and thus do not change the behavior of the variables in linear models. Statisticians standardize variables to make regression coefficients easier to interpret. Mean-centering at zero forces the intercept to 0,0, and allows one to interpret regression coefficients as the effect of one variable when all other variables are at their means.

For each dependent variable, we performed a two-step analysis. The first step was to calculate the zero-order correlation between each of the variables. These correlations represent how each demographic variable relates to the participation variables, at the most basic level, without considering interactive effects and intercorrelations between the demographic variables.

The second step consisted of running many ordinary least squares (OLS) regression models. The advantage of OLS regression over simple correlations is that it isolates the unique effects of each demographic variable, independent of the other variables with which it correlates. For example, low-income households are more likely to rent than higher-income households. Thus, the proportion of low-income households in a block group is correlated with the proportion of renters in a block group. Running a model that includes both renters and low-income tells us how the non-overlapping (or independent) portion of each of the demographic variables relates to participation. In other words, taking out the effects on participation that come from low income, how does renting relate to participation? And taking out the effects that renting has on participation, how does low income relate to participation?

We operated on a complete list of Massachusetts block groups that are served by PAs, after the trimming for block groups with very few PA customers (<25, which totals 33 block groups out of ~5000) and for which participation statistics were out of reasonable ranges. These block groups are considered the entire population, and the regression modeling was not attempting to make predictions for the future, so we ran the regressions with the entire population. We did not perform the usual training-test split but built our models on all observations (since values of the dependent variable for all of them were available for us).

The first step in the modeling was to create the model that explains the most amount of variance in the dependent variable (the "best-fit" model). To create the best-fit model, we used a forward stepwise regression technique. In this approach, we created a set of models with only one descriptive variable[7] each and chose one with the greatest R^2 as a starting point. Retaining this first descriptive variable in the model, we then added a second descriptive variable to the model and iterated through all the remaining descriptive variables to find the two-variable model with the greatest R^2. We continued the process

[7] These variables are also referred to as "independent variables" by statisticians.

of adding another descriptive variable until we got to the point where adding another descriptive variable did not result in a significant increase in R^2.

After the best-fit models, DNV ran several other models using combinations of variables that answer specific theory-based questions, such as how much of the overlapping effects of different variables are attributable to one or the other.

3.5 Individual modeling

3.5.1 Initial modeling

DNV attempted linear regression models on individual location-level data that were similar to the models used for the block group-level modeling described in the previous section. DNV implemented individual-level models for location participation (binary logistic model) and location savings rate (linear model) as dependent variables. The predictor variables were binary (0 or 1) dummy codes representing the different categorical data about individual locations. Those variables included:

- Pre- (or post-) 1950 construction as reported in the tax assessor data (2 levels)
- Renter (or owner), as newly determined in this study according to the methods described in Section 3.2 (2 levels)
- Several dummy codes to represent the different building size categories as newly determined in this study according to the methods described in Section 3.3 (6 levels)

Neither model produced interpretable results. The location model predicted that all records would be non-participants. The location savings rate model managed to explain less than 1% of the variance of the dependent variable.

After reconsidering that the predictor variables are all categorical variables and together represent 24 distinct categories, DNV determined that an ANOVA approach followed by a series of t-tests to test participation differences between individual pairs of categories would be a better statistical approach. Even if this approach failed to produce statistically significant differences, such an outcome would still be interpretable.

3.5.2 Data preparation

The individual models used the location-level participation rate as a binary variable, defined as 0 if the location did not ever participate between 2013 and 2019, and 1 if the location participated at least once during that period.

ANOVA analysis assumes that the dependent variable is normally distributed. DNV observed that the individual-level savings-rate variable was very positively skewed, so we developed a new variable with a distribution much closer to normal. This variable captures the depth of savings of an individual location *for all locations that participated.* This ratio was calculated as the sum of savings from 2013-2019 divided by the average consumption between 2013 and 2019. Years in which consumption was less than 1,000 kWh were excluded from the average calculation in the denominator.

$$Location\ Savings\ Rate = \frac{\sum_{2013}^{2019} \text{Savings}}{\text{mean}(\text{consumption})_{2013-2019}}$$

Where annual consumption <1000 kWh excluded from the denominator

To increase the normality of the distribution of this variable, DNV subjected the electric savings variable to a third-degree root transformation, and the gas savings to a sixth-degree root transformation.

3.5.3 New modeling

For the location participation models, DNV conducted a binary logistic regression using the binary location participation variable as the dependent variable, predicted by the combination of dummy variables that represented the 24 conditions (pre/post-1950 construction * renter/owner * 6 building size categories).

For the ANOVAs on the location savings rate, DNV observed that at the population level, the number of locations within the 24 individual conditions was not evenly distributed. To proceed with an ANOVA with an equal number of observations in each condition, DNV randomly selected 20 records for each condition for the electric analyses and 10 each for gas.[8] Then, to ensure that any findings were not due to the random selection process, DNV used a Monte Carlo approach where we repeated the random selection and all subsequent analyses 2,000 times. The ANOVA used the location savings rate described above as the dependent variable and the 24 categorical variables as predictors.

3.6 Provisions

3.6.1 Data preparation

The techniques used to add housing authority savings to the residential are known to undercount multifamily housing that has C&I accounts, such as by not identifying privately-owned low-income housing properties that may be on a C&I rate. Therefore, savings and consumption from multifamily housing are also undercounted. From a physics standpoint, consumption must be higher than savings for all uncounted locations, so statewide savings rates are likely slightly overstated. However, an individual block group-level multifamily building can represent most of the potential savings and consumption. Thus, the effects for individual block groups are varied and unknown for any block group.

The upstream lighting data used in this analysis comes from a geographic model of lighting savings and incentives developed by NMR. This geographic model was not available for 2013, so upstream lighting savings and incentives are not included in the location-level analysis for 2013.

The exclusion of behavioral and upstream data results in the following:

- Participation rates are understated relative to reality. It is highly probable that some households participated in an energy efficiency program via an in-store purchase (upstream) and are not reflected in the participation rate.
- A straight mathematical average of savings per location would be higher than reality because the savings would reflect the upstream sales but would not accurately reflect the number of locations that installed the lighting and generated the savings. Caution should be exercised for any population tables dividing savings by locations when including upstream sales data. This caveat does not apply to non-upstream savings tables.

3.6.2 Renter confidence

In general, we are confident that renters flagged by our algorithm are indeed renters. However, we are certain that there are housing authority and private subsidized housing locations that are underrepresented as renters.

3.6.3 Participation metrics

There is no perfect way to measure or define “participation.” This study utilized three different definitions, each of which has strengths and weaknesses.

- **Location participation** measures whether a building participated (yes/no). Location participation is best used to answer questions such as, “How many buildings did the program touch, regardless of size?” A strength of location participation is that it is stable over time, so it is useful for time-series analyses. Weaknesses of location participation are that it overstates participation of multifamily (MF) buildings because a location is considered a participant if any sub-unit within the building has participated; single-family (SF) homes count the same amount as MF buildings; and the building-level data grain does not match the ACS data grain, which is at the household level.
- **Population savings rate** measures the ratio of savings to consumption achieved in a geographic area. A strength of this metric is that locations with greater consumption (i.e., large MF buildings) have greater relative weight in the

[8] Ten observations per condition for gas produced more normal distributions than 20.

analyses. However, at a block group level, those large consumers can dominate block-group sums and averages while they do not have any exceptional weighting in individual account-level analyses. This can create disconnects between location participation rates and population savings rates in these block groups. Another limitation of this metric is that it does not fully account for consumption added for new construction.

- **Consumption-weighted location participation** uses location participation multiplied by the total consumption associated with each location. This metric gives a sense of how much of the underlying consumption the programs addressed. It is best used to account for the impact differences between SF and MF locations and to answer questions such as, “How well has the program addressed large customers?” The strengths of this variable are that it is stable over time, like location participation, and that it differentiates between SF and MF buildings. Weaknesses are that it overstates participation at MF locations to an even greater extent than location participation because of very large MF buildings; it is not at the same grain as ACS variables; and it can be difficult to interpret.

3.6.4 Model results

The models all had very low R^2 values, which means that there are a lot of factors that affect participation that the models are not able to account for. Among other limitations, models with low R^2 can appear to be unstable and produce inconsistent results because of all the unknown causes. R^2 can range from 0 to 1.0 and can be interpreted as how well the model predicts the independent variable. An R^2 of 1.0 means the model perfectly predicts the dependent variable—it accounts for 100% of the causal influences on the dependent variable. This is a theoretical value that can be approached but never fully achieved in practice—all models have some amount of error. In contrast, an R^2 of 0 means the variables in the model tell you nothing about the dependent variable. Our models achieved values between 0.2 and 0.1 which means that we are unable to account for 80 to 90% of the factors that influence participation.

The individual-level models failed to produce interpretable results.

4 ANALYSIS AND RESULTS

4.1 Consumption, participation, and savings summary

4.1.1 Time-series summaries of locations and consumption

Table 4-1 shows the number of unique physical locations for each PA and fuel for the period 2013 – 2019. Table 4-2 provides the total consumption of these physical locations for each PA and fuel for the period 2013–2017. The number of physical locations served by the PAs has increased somewhat over time; gas PAs may have more annual variability, as a physical location can move on or off gas service but remain habitable while most homes still require some form of grid-supplied electricity to be habitable. Critically, while the number of electric-served locations has increased by 7% from 2013 to 2019, electricity consumption has only increased by 1.6% over the same period. Savings gains from electric programs seem to be outpacing population growth. Part of this savings growth is due to a growing contribution from solar energy.

Table 4-1. Number of unique physical locations by year, 2013-2019

PA Name	Unique locations						
	2013	2014	2015	2016	2017	2018	2019
Cape Light Compact	120,380	129,307	130,110	129,352	129,859	130,452	170,676
Eversource	595,124	629,363	638,046	636,898	639,479	641,210	713,287
National Grid	742,279	798,811	800,945	798,023	799,894	800,580	870,160
Unitil	17,311	18,574	18,594	18,587	18,684	18,604	20,129
Electric total	**1,475,000**	**1,576,000**	**1,587,600**	**1,582,800**	**1,587,900**	**1,590,800**	**1,774,200**
Berkshire	24,825	27,421	25,613	27,754	28,054	27,970	32,165
EGMA	216,771	229,248	230,360	232,335	234,186	235,719	264,308
Eversource	182,043	175,275	176,591	177,386	178,504	179,741	203,462
Liberty	29,410	30,523	30,722	31,806	33,044	32,694	34,789
National Grid	539,531	579,050	579,969	579,515	583,794	587,524	637,230
Unitil	9,468	9,954	10,048	10,063	10,128	10,166	10,900
Gas total	**1,002,000**	**1,051,400**	**1,053,300**	**1,058,800**	**1,067,700**	**1,073,800**	**1,182,800**

*Additional views of all tables shown in this section are available on the MA Customer Profile interactive dashboard, located at https://insight.dnvgl.com/MACustomerProfile

Table 4-2. Annual consumption electric (MWh) and gas (Dth), 2013-2019

PA	Consumption (MWh, Dth)						
	2013	2014	2015	2016	2017	2018	2019
Cape Light Compact	1,087,934	1,095,535	1,138,019	1,117,069	1,082,640	1,163,346	1,144,691
Eversource	6,304,282	6,251,950	7,059,564	6,804,466	6,581,683	7,042,277	6,803,844
National Grid	8,292,336	8,739,224	8,736,781	8,209,122	7,976,179	8,528,990	8,015,238
Unitil	180,658	178,657	169,535	165,295	154,661	159,428	160,893
Electric total	**15,865,209**	**16,265,365**	**17,103,899**	**16,295,953**	**15,795,162**	**16,894,041**	**16,124,666**
Berkshire	2,730,163	3,281,635	3,259,136	2,676,099	2,995,378	3,731,839	3,307,917
EGMA	23,397,588	27,196,107	24,820,201	28,339,487	29,612,146	28,099,012	27,866,035
Eversource	18,931,151	21,081,252	24,221,658	20,411,584	21,339,967	24,071,983	24,172,021
Liberty	3,308,526	3,891,068	3,899,229	3,087,952	4,450,936	4,638,270	4,231,873
National Grid	58,710,894	71,482,920	71,377,137	61,925,903	65,197,063	73,374,931	72,343,036
Unitil	883,043	957,778	1,077,678	957,881	974,077	962,812	1,076,448
Gas total	**107,961,365**	**127,890,760**	**128,655,040**	**117,398,906**	**124,569,567**	**134,878,847**	**132,997,331**

4.1.2 Location-level participation and savings

Table 4-3 shows the statewide-level savings for upstream, behavioral, and all other programs for 2013-2019. Annual savings for electric programs have increased by 415,492 MWh (a 67% increase) between 2013 and 2019. The largest fraction of savings growth has been from upstream programs (an 85% increase), followed by other programs (a 55% increase), and behavioral programs (an 11% increase).

Annual savings for gas programs have increased by 474,096 Dth from 2013 through 2019 (a 29% increase). The largest fraction of savings growth has been from other programs (a 78% increase), followed by behavioral programs (a 9% increase), and upstream programs (no increase).

Table 4-3. Statewide savings for upstream, behavioral, and all other programs for 2013-2019.

Year	Electric savings (MWh)				Gas savings (Dth)			
	Upstream	Behavioral	Other	Total	Upstream	Behavioral	Other	Total
2013	397,372	109,776	113,478	**620,626**	438,427	533,737	675,981	**1,648,145**
2014	448,691	132,176	179,859	**760,726**	349,152	680,207	908,626	**1,937,985**
2015	324,787	134,333	204,922	**664,042**	329,061	672,718	903,753	**1,905,532**
2016	470,251	143,416	191,306	**804,973**	305,216	581,334	871,937	**1,758,487**
2017	618,587	136,589	178,339	**933,515**	366,202	599,916	870,776	**1,836,894**
2018	700,646	130,766	199,003	**1,030,415**	471,423	608,994	1,086,254	**2,166,671**
2019	738,117	121,980	176,021	**1,036,118**	334,194	587,908	1,200,112	**2,122,214**

Table 4-4 and Table 4-5 show the savings inclusive and exclusive of behavioral programs, respectively, for each PA in each year; these tables include upstream data. This analysis provides an important baseline for assessing multi-year participation, and for identifying trends in how participants engage in efficiency programs over time.

Upstream store sales data contributes a sizeable amount of savings to the PA totals, but as it is only captured at the store level, it cannot be linked to any individual physical location. This dynamic is likely to have the following effects on the results presented in this section:

- Participation rates represent the conservative floor relative to actual participation, as households that only participated in an energy efficiency program via an in-store purchase are not captured.
- A straight mathematical average of savings per location would be overstated as the savings reflects the upstream data, but not the accurate number of corresponding locations; caution should be exercised for any population-level tables dividing savings by locations. (This caveat does not apply to non-upstream participant-level tables.)
- Beyond upstream, participants through multifamily programs that were captured at a master meter level are not captured individually in the participant count; only the master-metered record is reflected.

Table 4-4. Participant savings (MWh, Dth) with behavioral and with upstream by year, 2013-2019

PA	Savings (MWh, Dth)							
	2013	2014	2015	2016	2017	2018	2019	Total Savings 2013-2019*
Cape Light Compact	19,752	23,538	22,308	29,269	42,306	56,307	49,214	796,859
Eversource	247,509	320,651	287,517	328,077	381,340	407,040	418,355	8,782,819
National Grid	350,279	413,121	350,181	443,780	506,363	564,437	563,827	11,668,491
Unitil	3,086	3,417	4,036	3,846	3,506	2,631	4,721	98,165
Electric (MWh) total	**620,626**	**760,726**	**664,042**	**804,973**	**933,515**	**1,030,415**	**1,036,118**	**21,346,334**
Berkshire	29,178	28,226	29,173	18,564	24,623	26,095	18,599	738,376
Columbia	191,842	205,669	252,223	265,581	241,290	306,372	337,811	6,574,769
Eversource	298,316	355,804	381,434	329,344	315,807	367,940	341,519	9,472,397
Liberty	30,957	30,003	19,987	21,865	20,302	27,736	25,509	725,727
National Grid	1,090,746	1,308,877	1,213,361	1,114,236	1,228,376	1,430,399	1,389,336	33,947,492
Unitil	7,106	9,407	9,355	8,898	6,496	8,130	9,703	234,007
Gas (Dth) total	**1,648,145**	**1,937,985**	**1,905,532**	**1,758,487**	**1,836,894**	**2,166,671**	**2,122,214**	**51,692,768**

*See Appendix B for an explanation of how total population savings are calculated.

Table 4-5. Participant savings (MWh, Dth) without behavioral and with upstream by year, 2013-2019

PA	Savings (MWh, Dth)							
	2013	2014	2015	2016	2017	2018	2019	Total Savings 2013-2019*
Cape Light Compact	19,752	23,538	22,308	29,269	42,306	56,307	49,369	797,013.53
Eversource	225,147	281,877	241,492	277,574	328,224	356,321	374,380	7,656,744.90
National Grid	262,864	319,718	261,873	350,868	422,890	484,390	486,475	9,195,117.31
Unitil	3,086	3,417	4,036	3,846	3,506	2,631	3,913	97,357.14
Electric (MWh) total	**510,850**	**628,550**	**529,709**	**661,557**	**796,926**	**899,649**	**914,138**	**17,746,232.89**
Berkshire	29,178	26,760	24,055	15,669	16,686	19,078	18,599	654,567.10
Columbia	191,842	205,669	252,223	265,581	241,290	306,372	337,811	6,574,769.30
Eversource	224,501	238,079	239,142	212,796	214,724	288,344	263,141	6,530,871.10
Liberty	30,957	30,003	19,987	21,865	20,302	27,736	25,509	725,996.50
National Grid	630,824	747,861	688,053	652,345	737,479	908,018	880,144	19,861,195.50
Unitil	7,106	9,407	9,355	8,898	6,496	8,130	9,102	233,399.50
Gas (Dth) total	**1,114,408**	**1,257,778**	**1,232,814**	**1,177,153**	**1,236,978**	**1,557,677**	**1,534,306**	**34,580,799.00**

*See Appendix B for an explanation of how total population savings are calculated.

Table 4-6 shows the savings exclusive of behavioral and upstream programs.

Table 4-6. Participant savings (MWh, Dth) without behavioral and without upstream by year, 2013-2019

PA	Savings (MWh, Dth)							
	2013	2014	2015	2016	2017	2018	2019	Total Savings 2013-2019*
Cape Light Compact	7,989	9,330	12,544	12,540	13,242	14,761	15,695	309,723
Eversource	57,592	78,857	95,918	83,224	77,286	84,291	75,160	2,164,373
National Grid	47,140	91,113	94,938	94,170	86,441	98,962	84,033	2,269,309
Unitil	757	560	1,522	1,372	1,370	989	1,133	28,973
Electric (MWh) total	**113,478**	**179,859**	**204,922**	**191,306**	**178,339**	**199,003**	**176,021**	**4,772,379**
Berkshire	16,705	16,113	15,576	10,691	12,837	13,497	11,157	410,914
Columbia	107,071	147,746	193,508	201,744	186,616	230,558	269,239	4,700,689
Eversource	160,827	183,488	188,984	159,207	161,470	193,437	193,064	4,872,816
Liberty	23,092	24,887	14,657	16,896	14,911	21,400	18,766	558,132
National Grid	363,832	528,704	483,660	476,293	489,633	620,109	701,689	13,453,325
Unitil	4,454	7,688	7,368	7,105	5,309	7,254	6,198	179,203
Gas (Dth) total	**675,981**	**908,626**	**903,753**	**871,937**	**870,776**	**1,086,254**	**1,200,112**	**24,175,079**

*See Appendix B for an explanation of how total population savings are calculated.

Table 4-7 shows the participation rates and population savings achieved by year and cumulatively for 2013-2017.

Table 4-7. Participation rates and population savings achieved by year, without behavioral and with upstream savings, 2013-2019

Year	Billing accounts	Billing locations	Consumption (MWh, Dth)	Tracking accounts	Tracking locations	Gross participant savings (MWh, Dth)	Account participation	Location participation	Population savings achieved
2013	2,300,126	1,457,742	15,865,209	173,047	126,087	510,850	7.5%	8.6%	3.2%
2014	2,677,898	1,558,093	16,265,365	185,357	155,805	628,550	6.9%	10.0%	3.9%
2015	2,837,893	1,571,017	17,103,899	174,237	148,604	529,709	6.1%	9.5%	3.1%
2016	2,629,027	1,564,278	16,295,953	168,933	141,901	661,557	6.4%	9.1%	4.1%
2017	2,653,241	1,569,254	15,795,162	183,079	152,876	797,680	6.9%	9.7%	5.1%
2018	2,767,654	1,572,225	16,894,041	219,170	205,465	899,650	7.9%	13.1%	5.3%
2019	2,841,333	1,754,119	16,124,666	212,552	208,932	914,138	7.5%	11.9%	5.7%
Electric (MWh) 2013-2019, cumulative	**4,627,904**	**1,855,614**	**114,344,296**	**1,059,539**	**832,228**	**4,942,133**	**22.9%**	**44.8%**	**NA**
2013	1,471,460	977,225	107,961,365	90,632	73,552	1,114,407	6.2%	7.5%	1.0%
2014	1,641,937	1,024,121	127,890,760	82,394	67,589	1,257,778	5.0%	6.6%	1.0%
2015	1,754,986	1,028,000	128,655,040	81,981	70,707	1,232,814	4.7%	6.9%	1.0%
2016	1,873,872	1,031,118	117,398,906	74,209	64,719	1,177,153	4.0%	6.3%	1.0%
2017	1,932,583	1,039,668	124,569,567	80,159	58,507	1,236,978	4.1%	5.6%	1.0%
2018	1,703,674	1,045,869	134,878,847	99,683	82,927	1,559,318	5.9%	7.9%	1.2%
2019	1,771,598	1,150,695	132,997,331	101,761	93,966	1,534,306	5.7%	8.2%	1.2%
Gas (Dth) 2013-2019, cumulative	**3,228,860**	**1,240,784**	**874,351,816**	**501,355**	**387,655**	**34,584,081**	**15.5%**	**31.2%**	**NA**

Table 4-8 below shows the number of new and repeat participating locations covering the 2013 and 2019 timeframe. Participants in this table exclude those who participated in upstream or behavioral programs. This data is excluded from this table for the following reasons. The behavioral program is a passive engagement offering. Customers are enrolled as either a treatment or a control and receive a mailing attempting to modify their behavior around energy use. Inclusion of the

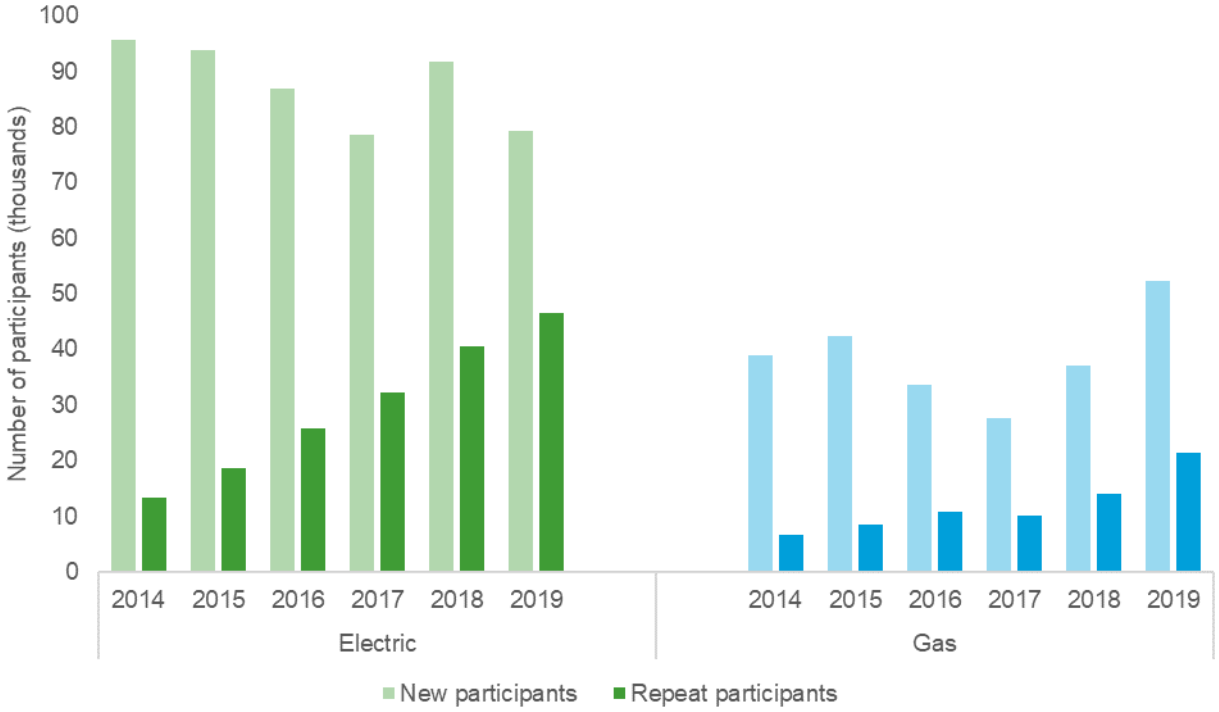
accounts in the behavioral program would cause participation rates for PAs that have a behavioral program to appear artificially high due to the passive participation. Inclusion of the behavioral savings would also make it difficult to have apples-to-apples comparisons between the PAs that do and do not offer behavioral programs. This data is reported at the point-of-sale (store) level and does not include any purchasing customer information; this means it is not possible to identify individual accounts, customers, or locations that participated in the program. The implication of this is that the upstream program does not impact any calculated participation rates in the customer profile report. In this table, new participants are those locations that participated in energy efficiency programs (exclusive of upstream and behavioral) in one year over the 2013 through 2019 timeframe. Repeat participants are those locations that participated in energy efficiency programs (exclusive of upstream and behavioral) in at least one prior year over the 2013 through 2019 timeframe. Of electric participants in 2019, 46,602 locations participated at least once from 2013 to 2019, and 79,185 locations did not participate between 2013 and 2019.

Table 4-8. New and repeat participating locations, without behavioral and without upstream

Participant type (locations)	2013	2014	2015	2016	2017	2018	2019
New participants	NA	95,684	93,715	86,878	78,522	91,647	79,185
Repeat participants	NA	13,424	18,590	25,836	32,347	40,504	46,602
Electric total	**86,725**	**109,108**	**112,305**	**112,714**	**110,869**	**132,151**	**125,787**
New participants	NA	38,897	42,372	33,606	27,634	37,013	52,328
Repeat participants	NA	6,714	8,481	10,756	10,021	14,035	21,507
Gas total	**50,255**	**45,611**	**50,853**	**44,362**	**37,655**	**51,048**	**73,835**

For both gas and electric, there is a positive trend in the number of repeat participants (see Figure 4-1). This suggests that the PAs are seeing growing numbers of customers returning to their program offerings year over year. The moderate apparent decrease in new participants seen for electric programs is an expected result of the programs maturing in the market. Note that these numbers exclude upstream and behavioral programs.

Figure 4-1. Number of new and repeat participating locations between 2014 and 2019 by fuel.



Annual unique location participation averaged around 9.4% for all PA-served locations (gas, electric, or combined gas and electric); this number represents a conservative floor, as it does not reflect either the behavioral participation or people that only participated in a PA program via the upstream store sales programs. Another point of note is that these tables reflect unique locations; multifamily locations and single-family locations are weighted equally in these tables. The location data grain is preferable for time series analysis due to its consistency across PAs and fuels, although a downside of location-level reporting is that it can obscure the number of people positively impacted by an efficiency measure. Electric participation rates averaged around 10.3%, and gas participation rates averaged around 7%. This may be at least partially reflective of electric customers having more diversity in the energy-consuming measures in their locations to upgrade via energy efficiency programs—particularly appliances (e.g., lighting, TVs, computers) for which the upgrade cost for energy efficiency is not on the same scale as replacing a heating system or insulating a house. Additional PA-specific information can be found on the MA Customer Profile Dashboard. Tables Table 4-9, Table 4-10, and Table 4-11 summarize participation rates between gas and electric combined, electric-only, and gas-only, respectively.

Table 4-9. Time-series participation rates for unique electric and/or gas PA-served locations, without upstream or behavioral.

Year	Combined electric and gas billing locations	Combined electric and gas tracking locations	Unique location participation rate
2013	1,637,541	138,651	8.5%
2014	1,725,217	133,197	7.7%
2015	1,733,167	138,390	8.0%
2016	1,741,733	145,131	8.3%
2017	1,746,849	143,143	8.2%
2018	1,749,347	178,866	10.2%
2019	1,948,611	178,367	9.2%
2013 - 2019 Cumulative	**2,062,470**	**821,267**	**39.8%**

Table 4-10. Time-series participation rates for unique electric PA-served locations, without upstream or behavioral.

Year	Electric billing locations	Electric tracking locations	Unique location participation rate
2013	1,457,742	88,922	6.1%
2014	1,558,093	109,266	7.0%
2015	1,571,017	112,443	7.2%
2016	1,564,278	111,393	7.1%
2017	1,569,254	109,056	6.9%
2018	1,572,225	139,931	8.9%
2019	1,754,119	131,000	7.5%
2013 - 2019 Cumulative	**1,855,614**	**630,765**	**34.0%**

Table 4-11. Time-series participation rates for unique gas PA-served locations, without upstream or behavioral.

Year	Gas billing locations	Gas tracking locations	Unique location participation rate
2013	977,225	50,570	5.2%
2014	1,024,121	45,731	4.5%
2015	1,028,000	50,885	4.9%
2016	1,031,118	44,315	4.3%
2017	1,039,668	33,806	3.3%
2018	1,045,869	46,919	4.5%
2019	1,150,695	61,782	5.4%
2013 - 2019 Cumulative	**1,240,784**	**272,016**	**21.9%**

Table 4-12 - Table 4-14 present participation rates and population savings achieved by PA from 2013-2019. DNV did not include behavioral savings in these population-level tables. This was done to ensure that the savings reported reflected initiatives that required active customer participation decisions, and because behavioral initiatives are not offered in all service territories. Consistent with prior reporting, upstream lighting savings are included in these tables but are not reflected in participant counts.

Table 4-12. Participation rates by PA without behavioral and without upstream, cumulative from 2013

PA	Billing accounts	Billing locations	Tracking accounts	Tracking locations	Account participation	Location participation
Cape Light Compact	289,277	177,665	54,685	44,876	18.9%	25.3%
Eversource	2,227,243	744,785	355,733	266,838	16.0%	35.8%
National Grid	2,056,117	912,270	387,379	313,859	18.8%	34.4%
Unitil	55,267	20,894	5,941	5,190	10.7%	24.8%
Electric total	**4,627,904**	**1,855,614**	**803,738**	**630,763**	**17.4%**	**34.0%**
Berkshire	66,385	33,631	7,472	5,785	11.3%	17.2%
Columbia	912,732	270,245	73,497	40,985	8.1%	15.2%
Eversource	626,856	222,097	68,523	56,410	10.9%	25.4%
Liberty	133,199	38,087	6,617	4,978	5.0%	13.1%
National Grid	1,455,345	665,368	195,072	162,598	13.4%	24.4%
Unitil	34,343	11,356	1,842	1,650	5.4%	14.5%
Gas total	**3,228,860**	**1,240,784**	**353,023**	**272,406**	**10.9%**	**22.0%**

Table 4-13. Participation rates and population savings achieved by PA, without behavioral and with upstream savings, 2019

PA	Billing accounts	Billing locations	Consumption (MWh, Dth)	Tracking accounts	Tracking locations	Gross participant savings (MWh, Dth)	Account participation	Location participation	Population savings achieved
Cape Light Compact	189,595	170,677	1,144,691	17,929	18,100	49,369	9.5%	10.6%	4.3%
Eversource	1,282,168	713,270	6,803,844	85,200	83,037	374,380	6.6%	11.6%	5.5%
National Grid	1,338,353	870,172	8,015,238	107,258	105,644	486,475	8.0%	12.1%	6.1%
Unitil	31,217	20,129	160,893	2,165	2,151	3,913	6.9%	10.7%	2.4%
Electric (MWh) total	**2,841,333**	**1,774,248**	**16,124,666**	**212,552**	**208,932**	**914,138**	**7.5%**	**11.8%**	**5.7%**
Berkshire	44,499	32,170	3,307,917	1,471	1,366	18,599	3.3%	4.2%	0.6%
Columbia	346,763	264,307	27,866,035	22,680	21,088	337,811	6.5%	8.0%	1.2%
Eversource	326,185	203,463	24,172,021	20,210	18,735	263,141	6.2%	9.2%	1.1%
Liberty	62,129	34,789	4,231,873	1,852	1,461	25,509	3.0%	4.2%	0.6%
National Grid	973,772	637,236	72,343,036	54,925	50,737	880,144	5.6%	8.0%	1.2%
Unitil	18,250	10,900	1,076,448	623	579	9,102	3.4%	5.3%	0.8%
Gas (Dth) total	**1,771,598**	**1,182,865**	**1,329,973,308**	**101,761**	**93,966**	**1,534,306**	**5.7%**	**7.9%**	**1.2%**

Table 4-14. Participation rates and population savings achieved by PA, without behavioral and with upstream savings, 2018

PA	Billing accounts	Billing locations	Consumption (MWh, Dth)	Tracking accounts	Tracking locations	Gross participant savings (MWh, Dth)	Account participation	Location participation	Population savings achieved
Cape Light Compact	192,276	130,452	1,163,346	16,670	16,198	56,307	8.7%	12.4%	4.8%
Eversource	1,269,421	641,210	7,042,277	94,366	86,316	356,321	7.4%	13.5%	5.1%
National Grid	1,278,758	800,580	8,528,990	106,360	101,215	484,390	8.3%	12.6%	5.7%
Unitil	27,199	18,604	159,428	1,773	1,736	2,632	6.5%	9.3%	1.7%
Electric (MWh) total	**2,767,654**	**1,590,846**	**16,894,041**	**219,169**	**205,465**	**899,650**	**7.9%**	**12.9%**	**5.3%**
Berkshire	35,507	27,970	3,731,839	1,560	1,504	19,078	4.4%	5.4%	0.5%
Columbia	343,847	235,719	28,099,012	22,088	8,625	306,372	6.4%	3.7%	1.1%
Eversource	322,163	179,741	24,071,983	19,736	19,656	288,344	6.1%	10.9%	1.2%
Liberty	62,512	32,694	4,638,270	1,644	1,610	27,736	2.6%	4.9%	0.6%
National Grid	924,145	587,524	73,374,931	54,163	51,079	908,018	5.9%	8.7%	1.2%
Unitil	15,500	10,166	962,812	492	453	9,771	3.2%	4.5%	1.0%
Gas (Dth) total	**1,703,674**	**1,073,814**	**134,878,847**	**99,683**	**82,927**	**1,559,318**	**5.9%**	**7.7%**	**1.2%**

4.1.3 Program-level participation and savings

Table 4-15 - Figure 4-19 show the number of participating locations and saving by program between 2013 and 2019. Residential existing buildings continue to achieve the highest proportion of overall savings, both with and without upstream savings included. Savings from all programs, and across both fuels, have followed an increasing trend since 2013, except for savings from non-behavioral, non-upstream electric, which increased from 2013-2015, then remained steady or declined slightly (see Figure 4-2)

Table 4-15. Participant locations served by program, without behavioral and without upstream, 2013-2019

Program	Locations served							
	2013	2014	2015	2016	2017	2018	2019	2013-2019
Residential New Homes	200	1,751	3,005	3,062	3,184	4,852	4,829	20,883
Residential Existing Buildings	110,028	142,274	135,055	127,036	135,551	186,776	189,801	1,026,521
Income Eligible Existing Buildings	12,527	13,102	12,979	13,052	15,557	14,707	15,384	97,308
Electric total	**122,755**	**157,127**	**151,039**	**143,150**	**154,292**	**206,335**	**210,014**	1,144,712
Residential New Homes	100	973	1,039	2,045	2,091	3,075	2,939	**12,262**
Residential Existing Buildings	68,529	64,594	67,409	60,777	52,450	76,414	88,300	**478,473**
Income Eligible Existing Buildings	2,646	2,077	2,219	2,094	1,739	3,493	3,158	**17,426**
Gas total	**71,275**	**67,644**	**70,667**	**64,916**	**56,280**	**82,982**	**94,397**	**508,161**

Table 4-16. Population savings by program, without behavioral and with upstream, 2013-2019

Program	Savings (MWh, Dth)							
	2013	2014	2015	2016	2017	2018	2019	2013-2019*
Residential New Homes	2,832	12,919	11,047	16,470	18,453	19,390	21,309	333,900
Residential Existing Buildings	478,146	571,475	473,494	603,952	735,437	831,577	856,085	16,284,691
Income Eligible Existing Buildings	29,872	44,157	45,168	41,135	43,790	48,692	36,745	1,129,922
Electric (MWh) total	**510,850**	**628,550**	**529,709**	**661,557**	**797,680**	**899,659**	**914,138**	17,748,513
Residential New Homes	58,312	93,494	113,367	84,733	116,252	107,226	97,703	**2,535,827**
Residential Existing Buildings	893,954	898,739	887,934	873,657	878,309	1,201,739	1,163,367	**25,786,180**
Income Eligible Existing Buildings	162,142	265,545	231,513	218,763	242,417	250,354	273,236	**6,262,074**
Gas (Dth) total	**1,114,407**	**1,257,778**	**1,232,814**	**1,177,153**	**1,236,978**	**1,559,318**	**1,534,306**	**34,584,080**

Table 4-17. Population savings by program, without behavioral and without upstream, 2013-2019

Program	Savings (MWh, Dth)							
	2013	2014	2015	2016	2017	2018	2019	2013-2019*
Residential New Homes	2,832	12,919	11,047	16,470	18,453	19,390	21,309	333,900.16
Residential Existing Buildings	80,774	122,783	148,707	133,701	116,849	130,921	117,968	3,310,816.05
Income Eligible Existing Buildings	29,872	44,157	45,168	41,135	43,790	48,692	36,745	1,129,921.87
Electric (MWh) total	**113,906**	**179,859**	**204,923**	**191,393**	**179,092**	**199,106**	**176,266**	**4,774,638**
Residential New Homes	58,312	93,494	113,367	84,733	116,252	107,226	97,703	25,358,265.71
Residential Existing Buildings	455,527	549,587	558,873	568,441	512,107	728,721	829,172	176,892,742.18
Income Eligible Existing Buildings	162,142	265,545	231,513	218,763	242,417	250,354	273,236	62,620,739.81
Gas (Dth) total	**675,980**	**908,626**	**903,753**	**871,937**	**870,776**	**1,086,300**	**1,200,112**	**264,871,748**

Figure 4-2. Trend for electric population savings across Residential New Homes, Residential Existing Buildings, and Income Eligible Existing Buildings from 2013-2019

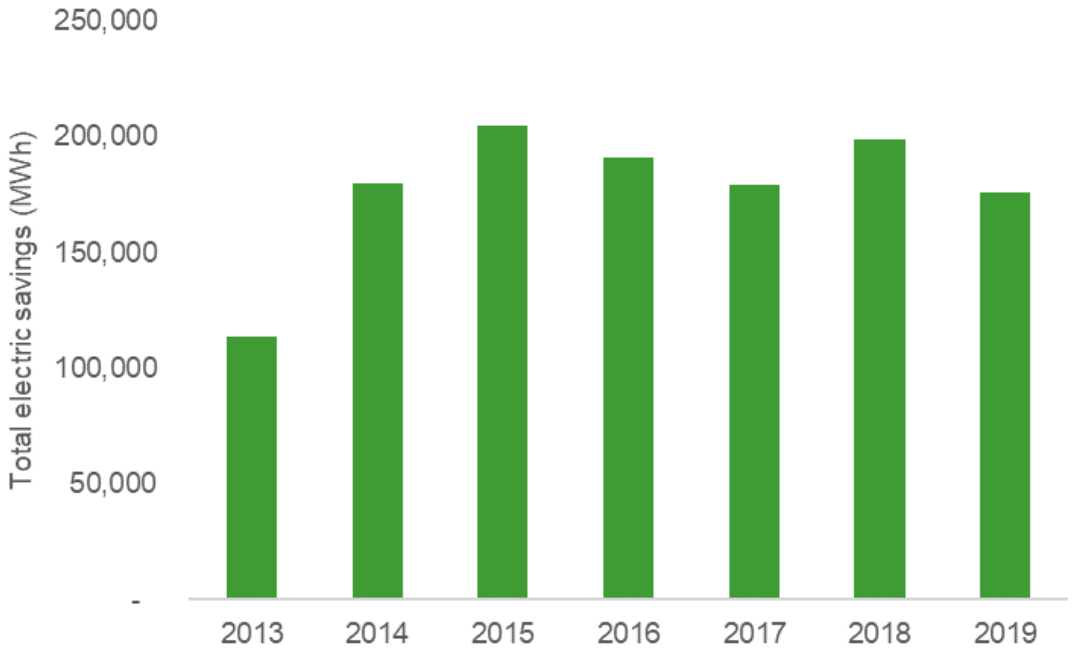


Table 4-18. Population savings by PA and program without behavioral and with upstream, 2019

PA	Income Eligible Existing Buildings		Residential New Buildings		Residential Existing Buildings	
	Savings (MWh, Dth)	Percentage of savings	Savings (MWh, Dth)	Percentage of savings	Savings (MWh, Dth)	Percentage of savings
Cape Light Compact	2,090	4.2%	1,604	3.2%	45,674	92.5%
Eversource	18,609	5.0%	11,363	3.0%	344,409	92.0%
National Grid	15,643	3.2%	8,264	1.7%	462,568	95.1%
Unitil	402	10.3%	78	2.0%	3,434	87.7%
Electric (MWh) total	**36,745**	**4.0%**	**21,309**	**2.3%**	**856,085**	**93.6%**
Berkshire	2,996	16.1%	767	4.1%	14,837	79.8%
Columbia	40,896	12.1%	14,581	4.3%	282,334	83.6%
Eversource	44,295	16.8%	21,406	8.1%	197,440	75.0%
Liberty	5,881	23.1%	1,984	7.8%	17,644	69.2%
National Grid	177,202	20.1%	58,615	6.7%	644,327	73.2%
Unitil	1,966	21.6%	350	3.8%	6,786	74.6%
Gas (Dth) total	**273,236**	**17.8%**	**97,703**	**6.4%**	**1,163,367**	**75.8%**

Table 4-19. Population savings by PA and program without behavioral and with upstream, 2018

PA	Low-Income Whole House		Residential Whole House		Residential Products	
	Savings (MWh, Dth)	Percentage of savings	Savings (MWh, Dth)	Percentage of savings	Savings (MWh, Dth)	Percentage of savings
Cape Light Compact	2,036	3.6%	1,045	1.9%	53,225	94.5%
Eversource	22,927	6.4%	9,978	2.8%	323,414	90.8%
National Grid	23,472	4.8%	8,326	1.7%	452,594	93.4%
Unitil	256	9.7%	41	1.6%	2,344	88.7%
Electric (MWh) total	**48,692**	**5.4%**	**19,390**	**2.2%**	**831,577**	**92.4%**
Berkshire	4,334	22.7%	576	3.0%	14,168	74.3%
Columbia	37,498	12.2%	14,401	4.7%	254,474	83.1%
Eversource	42,958	14.9%	26,567	9.2%	218,818	75.9%
Liberty	5,583	20.1%	1,543	5.6%	20,609	74.3%
National Grid	158,590	17.5%	63,841	7.0%	685,588	75.5%
Unitil	1,390	14.2%	298	3.1%	8,083	82.7%
Gas (Dth) total	**250,354**	**16.1%**	**107,226**	**6.9%**	**1,201,739**	**77.1%**

4.2 End-use analysis

This section presents statewide-level summaries of savings and incentives across 10 end-use categories (7 electric measures and 3 gas measures). We also split these categories across delivery channels: downstream, upstream, direct install, and income-eligible direct install. The purpose of this analysis is to offer insight into trends in these metrics over time. The measures shown in this section were chosen because they represent the highest fraction of savings for each fuel type. Additional information about trends in these measure categories is available on the interactive MA Customer Profile Dashboard. Examples of measures included in these broader end use categories are found in Table 4-20.

Table 4-20. Examples of measures included in broad end use categories.

Fuel	End use category	Measures
Gas and Electric	Envelope	Air sealing
		Weatherization
		Insulation
	HVAC	Heating systems
		Duct sealing/insulation
		Thermostats
	Hot Water	Faucet Aerators
		Showerheads
		Water heaters
Electric	Lighting	Bulbs
		Fixtures
		Sensors
	Plug Load	Smart strips

Fuel	End use category	Measures
	TLC Kits	Bundled measures including, e.g., aerators, LED night lights, thermometers
	Appliance	Consumer products like refrigerators

4.2.1 Electric end uses

Figure 4-3 shows a comparison between the share of savings and the share of cost between each electric incentive. Although total savings varies annually, these charts are normalized to 100% to facilitate comparisons of the proportions of savings and incentives by measure type. The most notable trend in these figures is that lighting accounts for most incentives and savings. However, the proportion of savings accounted for by lighting is much greater than the proportion of incentives. This pattern indicates that lighting has relatively high benefit-cost ratios. The pattern is reversed for envelope and HVAC measures, which indicate relatively low benefit-cost ratios.

Figure 4-3. Comparison between savings and incentives contribution for electric measures

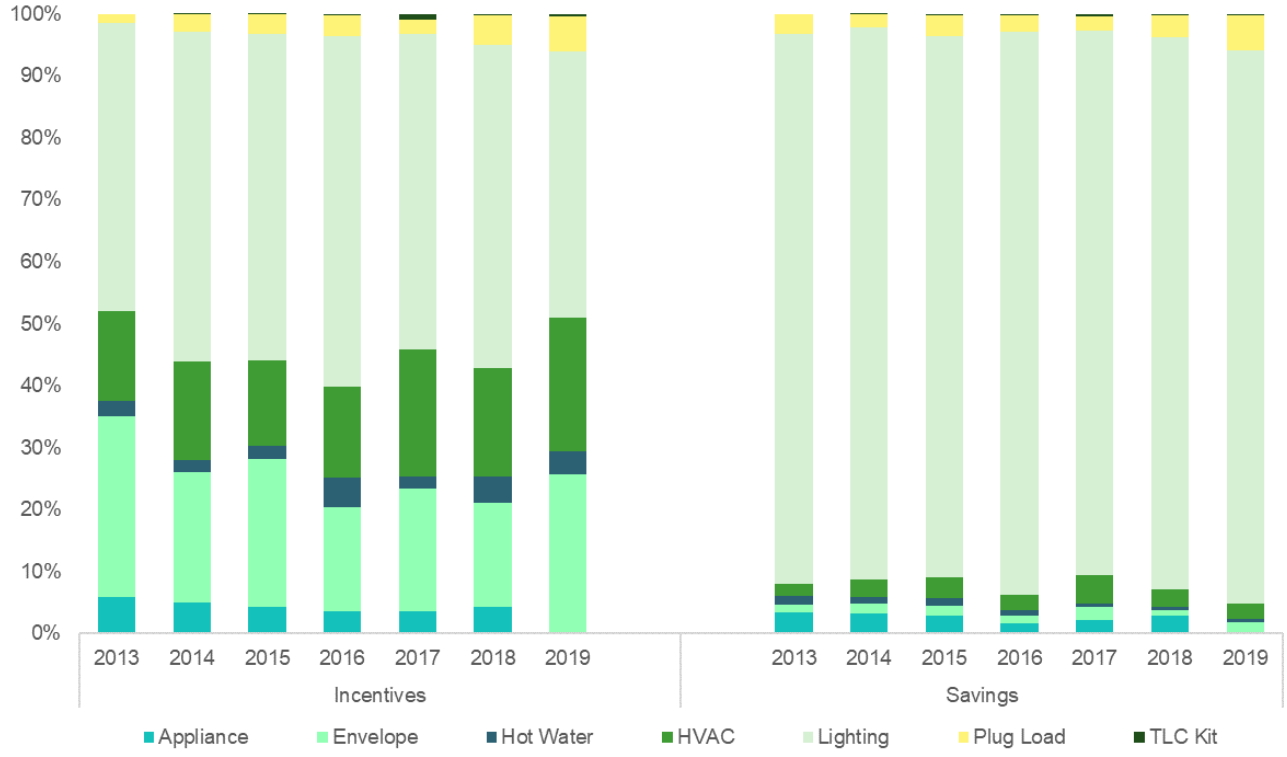


Figure 4-4 shows the distribution of electric incentives and savings across the four delivery channels: direct install, downstream, income-eligible direct install, and upstream. The upstream program accounts for most savings and a much smaller proportion of incentives. This indicates that upstream measures have a relatively high benefits/costs ratio. Most upstream electric measures are lighting, so this is consistent with the previous figures. In contrast, the other delivery channels each account for a greater proportion of the incentives than savings. This indicates relatively low benefits/costs ratios. Income eligible direct install has the worst favorable ratio in these figures.

Figure 4-4. Comparison between savings and incentives contribution for electric delivery pathways

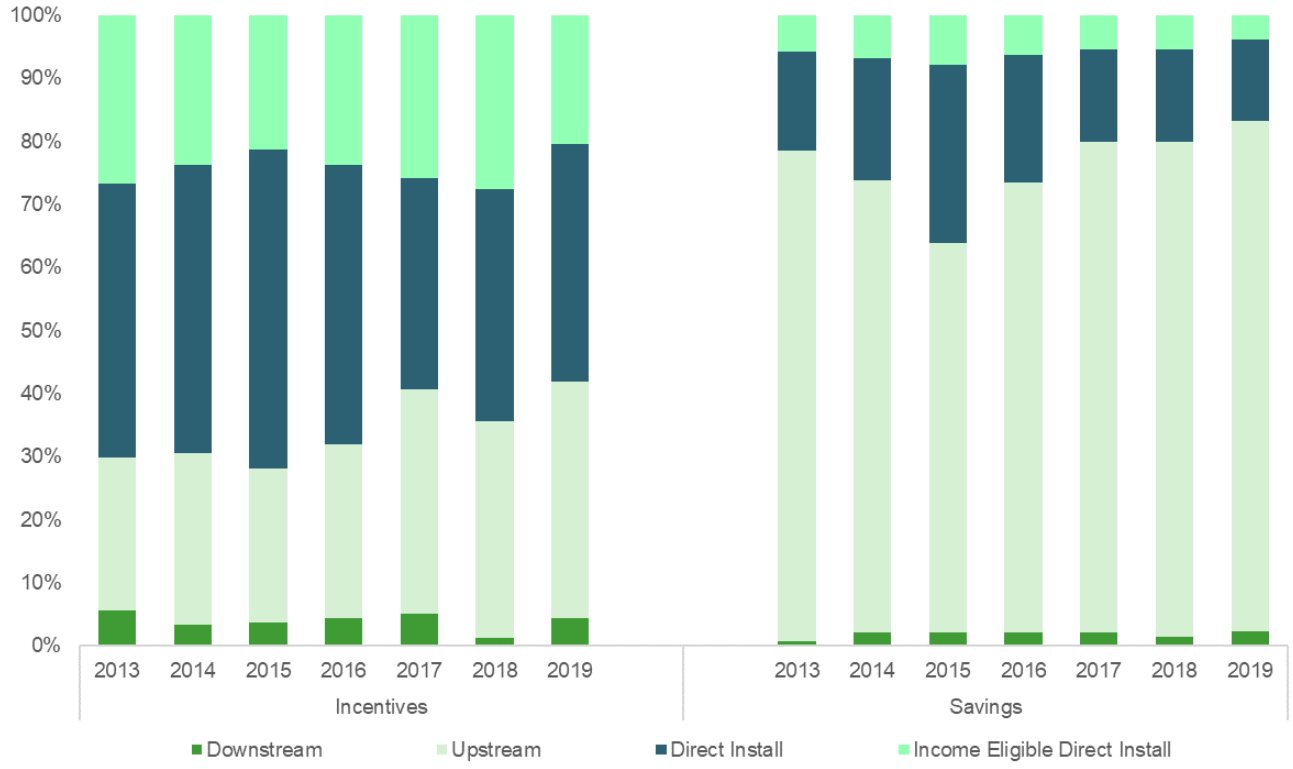
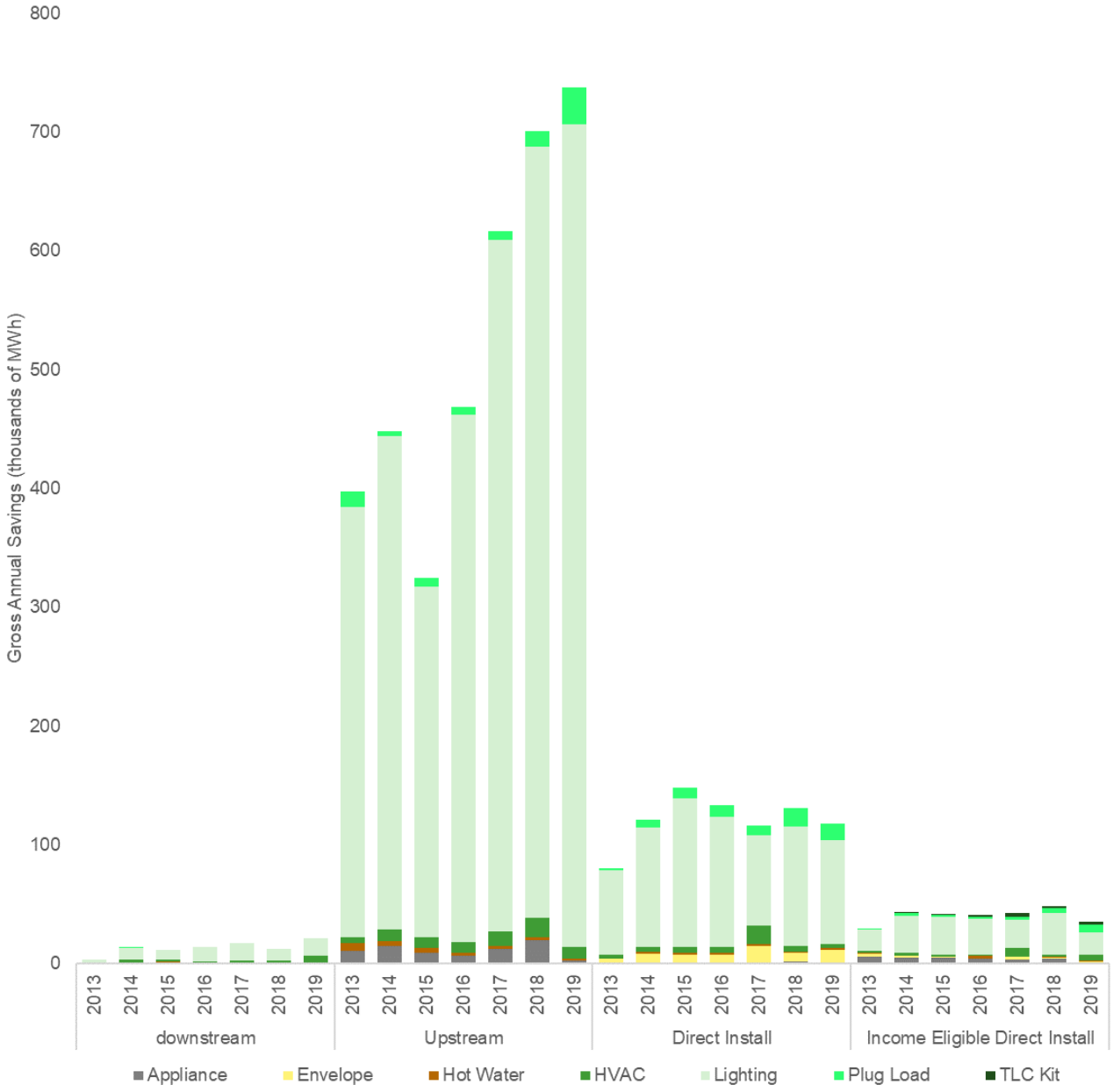


Figure 4-5 combines the savings information presented in the two previous figures. This presentation makes it easier to see that the upstream channel has experienced the most growth since 2013, and that almost all the savings come from lighting measures.

Figure 4-5. Electric savings by year and channel



The most important implication of these patterns is that the PAs are going to find it more difficult to achieve overall program savings goals *and* pass cost-effectiveness tests when lighting becomes ineligible sometime in the early or mid-2020s.

4.2.2 Gas end uses

Figure 4-6 shows a similar comparison of the proportion of incentives and savings for the gas measures. For gas measures, the proportions are more evenly distributed among the measure types and the proportions of incentives and savings are more consistent for each measure type. There appears to be a trend toward HVAC representing smaller proportions of incentives than savings, and vice-versa for envelope measures. This suggests the benefit-cost ratios for HVAC are slightly higher than those for envelope.

Figure 4-6. Comparison between savings and incentives contribution for gas measures

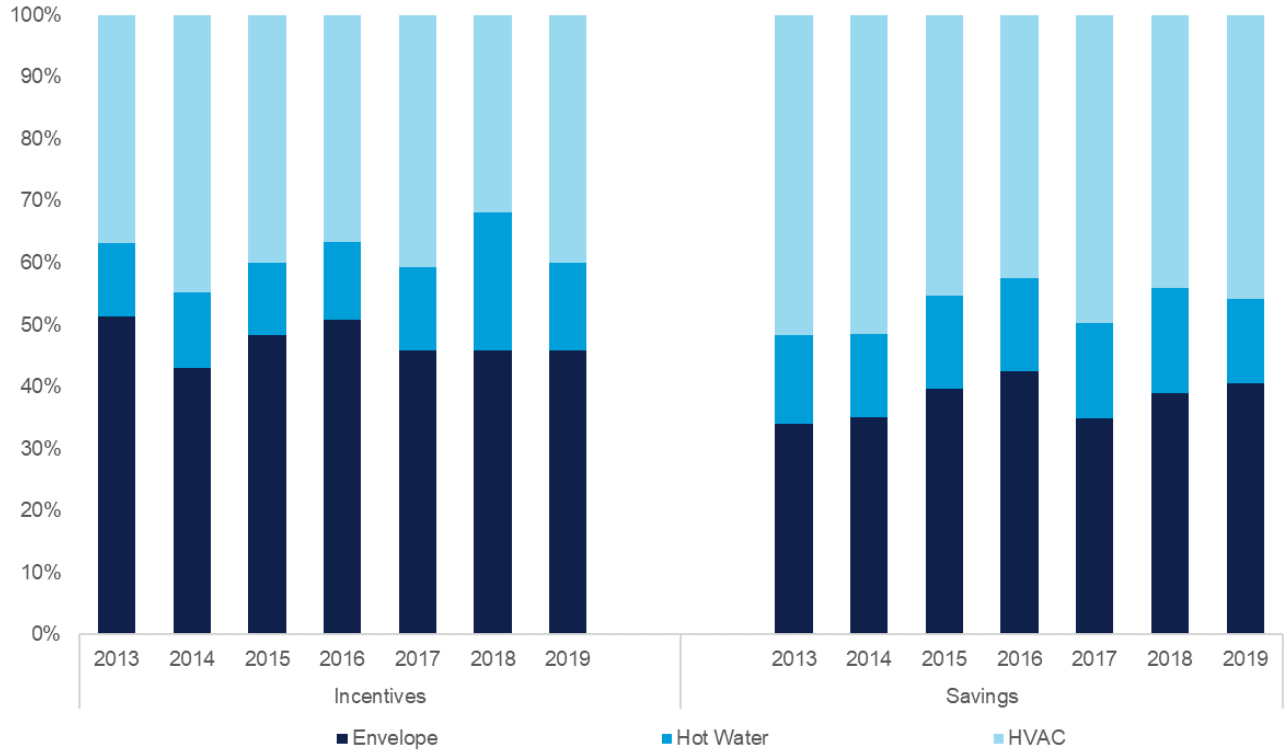


Figure 4-7 shows the distribution of gas incentives and savings across the four delivery channels: direct install, downstream, income-eligible direct install, and upstream. The most apparent trend is that direct install has slowly and steadily accounted for an increasing proportion of the gas incentives and savings. This program share seems to be coming primarily from the upstream channel, which displays a slightly decreasing trend.

Figure 4-7. Comparison between savings and incentives contribution for gas delivery pathways

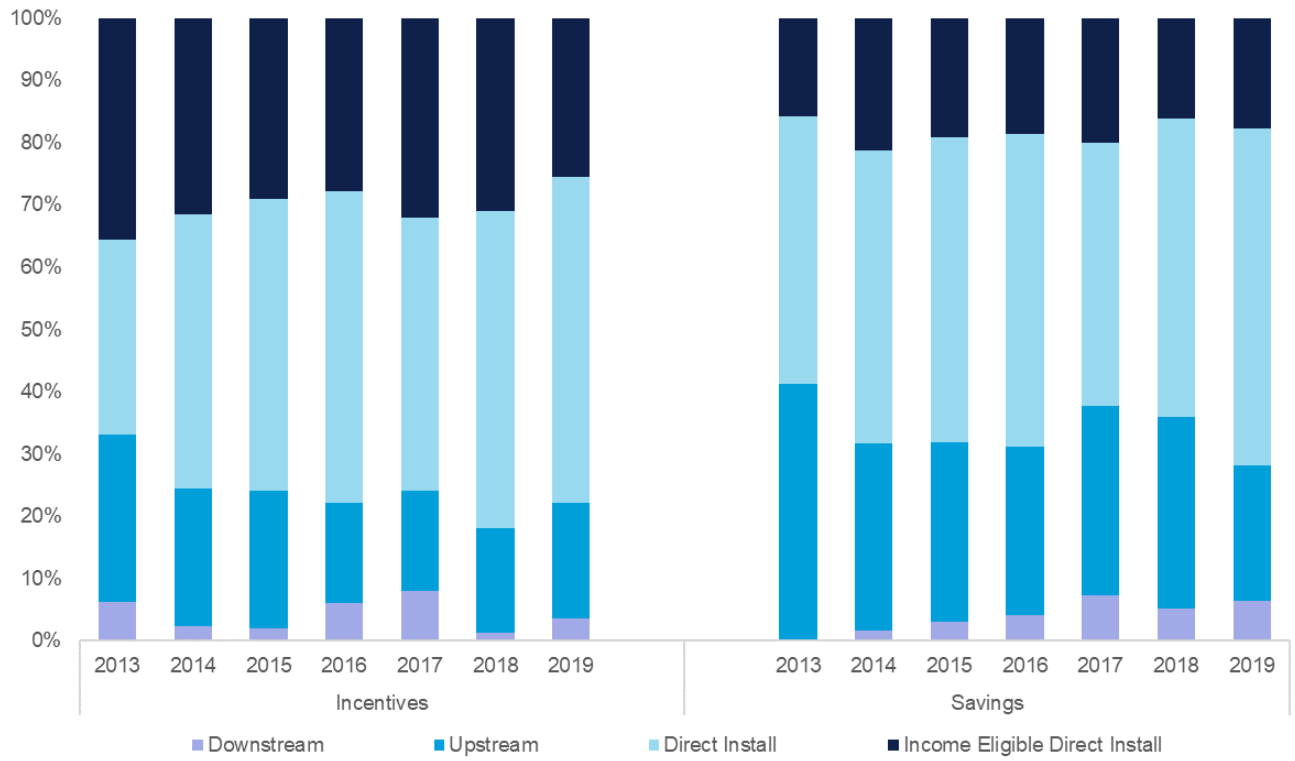
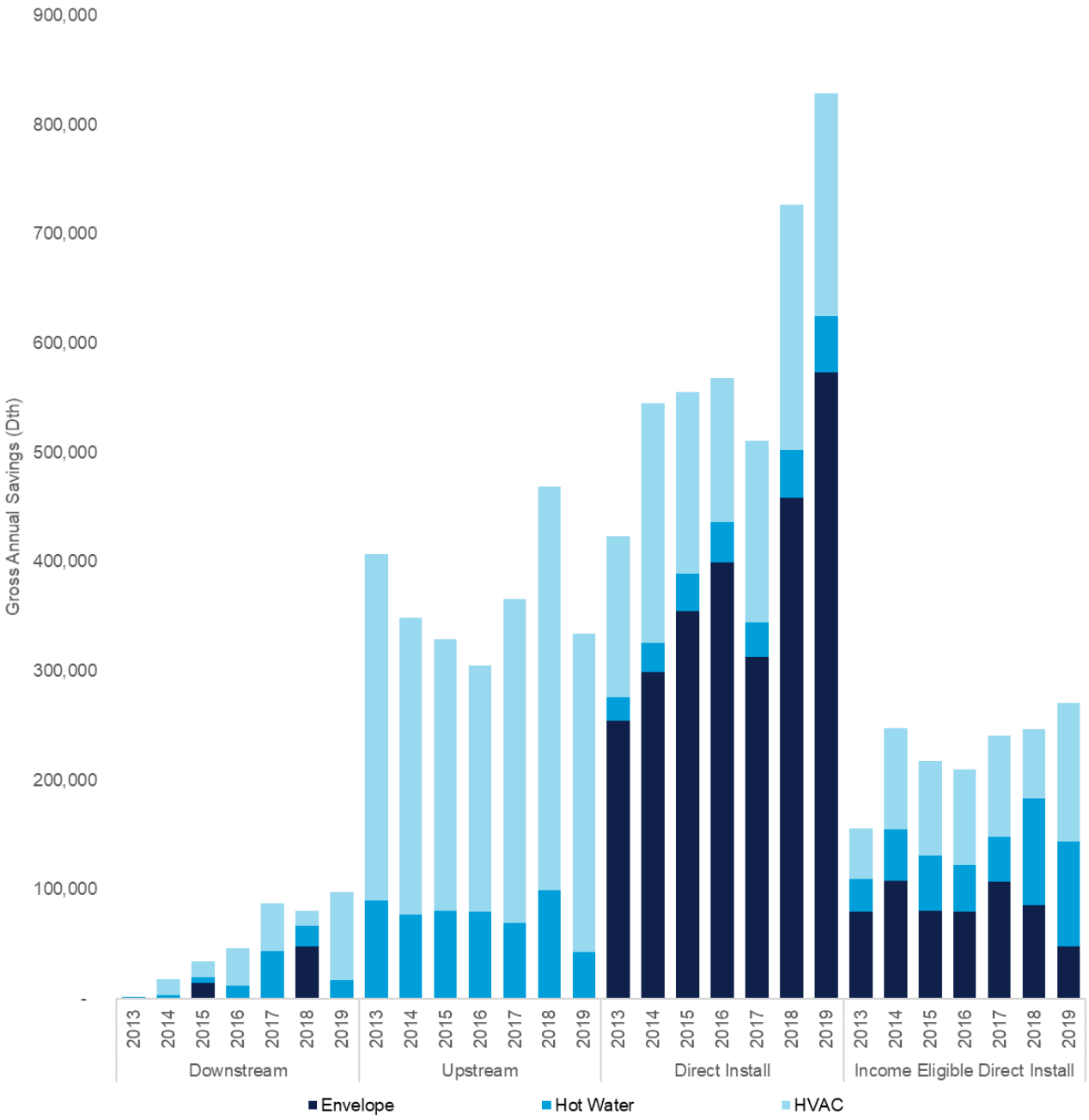


Figure 4-8 shows the combination of the two charts above for gas savings. Presented this way, the growth trend for the direct install program is more apparent, while any trend for the upstream program is difficult to ascertain against year-to-year variability. Additionally, a slight growth trend is apparent for both the downstream and income-eligible delivery channels. Another trend this figure illuminates is that most of the growth is accounted for by envelope measures.

For downstream, direct install, and income-eligible direct install pathways, there is a trend of overall increasing savings. There are no clear measure-level trends in downstream. For direct install, there is a clear increasing trend in envelope and hot water end uses. For income-eligible direct install, hot water has a visible increasing trend. For upstream, there is no strong trend overall; the measure mix appears to be steady year to year as well.

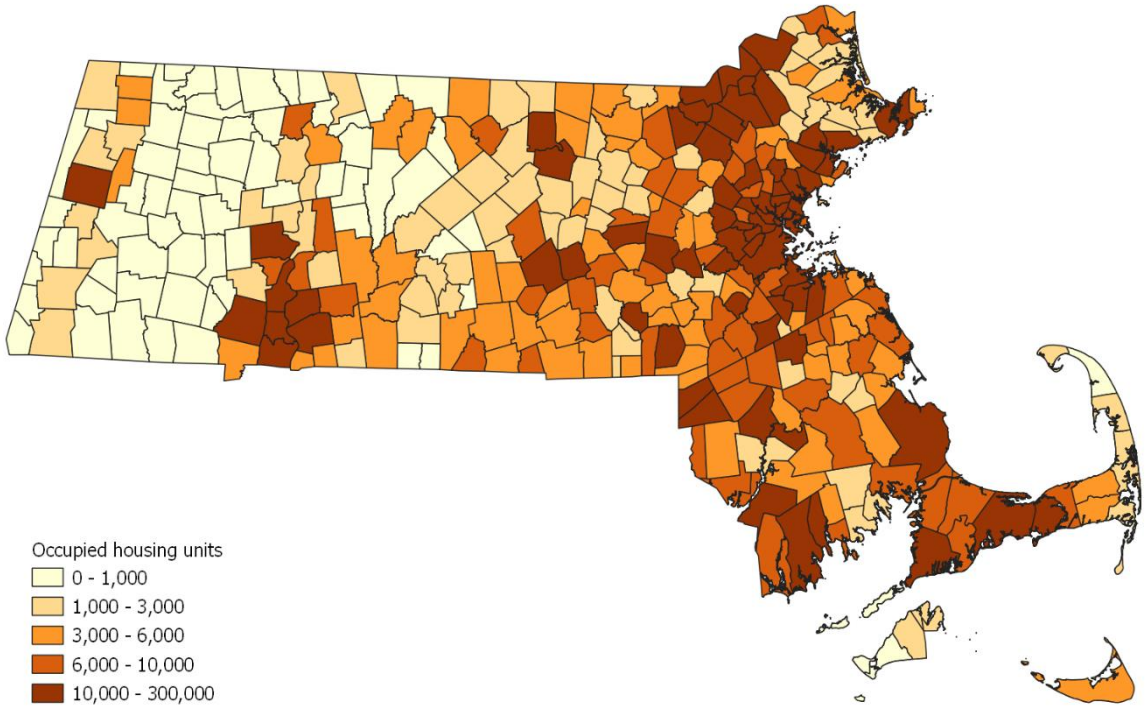
Figure 4-8. Gas savings by year and channel



4.3 Geographic market analysis

This section provides geographic summaries of program incentives relative to town-level consumption. This summarization is done by dividing the town-level incentives by the consumption for the entire block group. The results have been mapped to show how incentives vary geographically. To protect customer privacy, some data is masked where there are a minimum number of premises (fewer than 100 premises). Some towns are served by municipal electric or gas utilities rather than the PAs, so for these towns, no consumption-normalized incentive data is available. Figure 4-9 shows the number of occupied housing units in each town across the state, taken from the ACS 5-year averages.

Figure 4-9. Population size represented by the number of occupied housing units by town



4.3.1 Electric geographic analysis

Figure 4-10 shows incentive dollars spent cumulatively in 2013 through 2019 per kWh of consumption, aggregated to the town level. Lighter green areas of the map show the areas with the lowest dollars spent per unit of energy usage, while darker green areas show regions of higher spending. Notably, there is no apparent straightforward relationship between population size and consumption-weighted spending. In these maps, a model was used to geographically allocate upstream lighting incentives to the block group level, and these were aggregated to the town level to indicate upstream incentives.

Figure 4-10. Map of incentive dollars per kWh consumption by town, cumulative 2013-2019

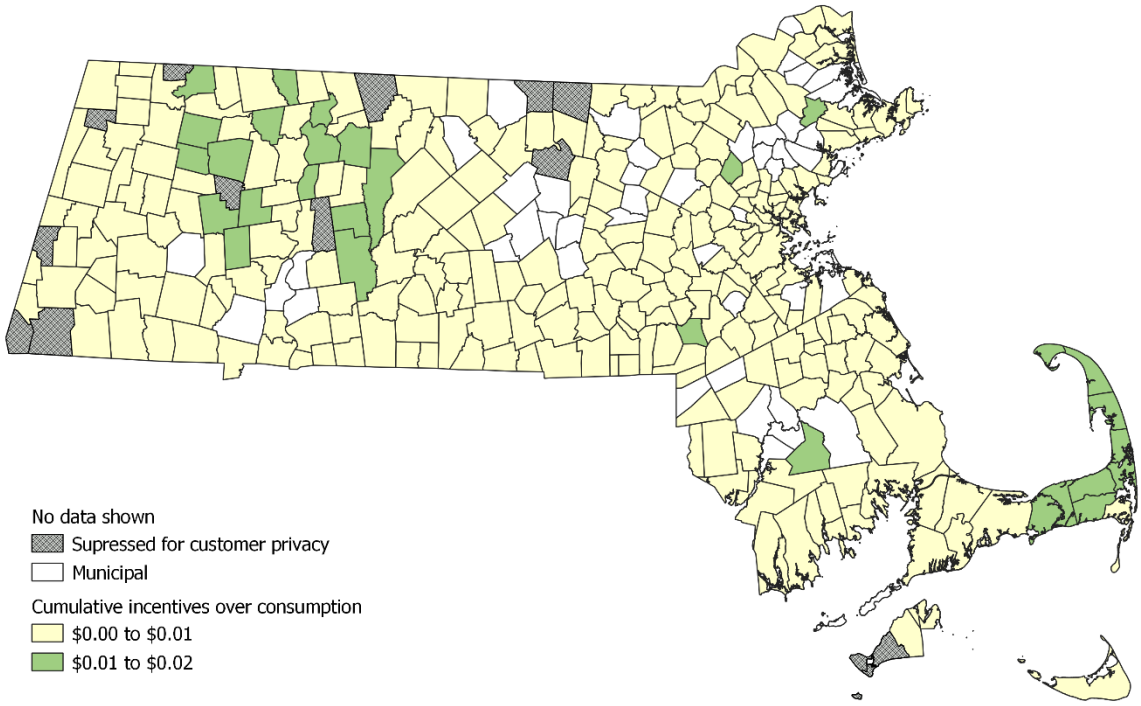


Figure 4-11 and Figure 4-12 show the incentive dollars spent per kWh of consumption in 2018 and 2019, respectively. While there is some year-over-year variability, the overall pattern between years remains consistent with the longer-term view shown in Figure 4-10.

Figure 4-11. Map of incentive dollars per kWh consumption by town, 2018

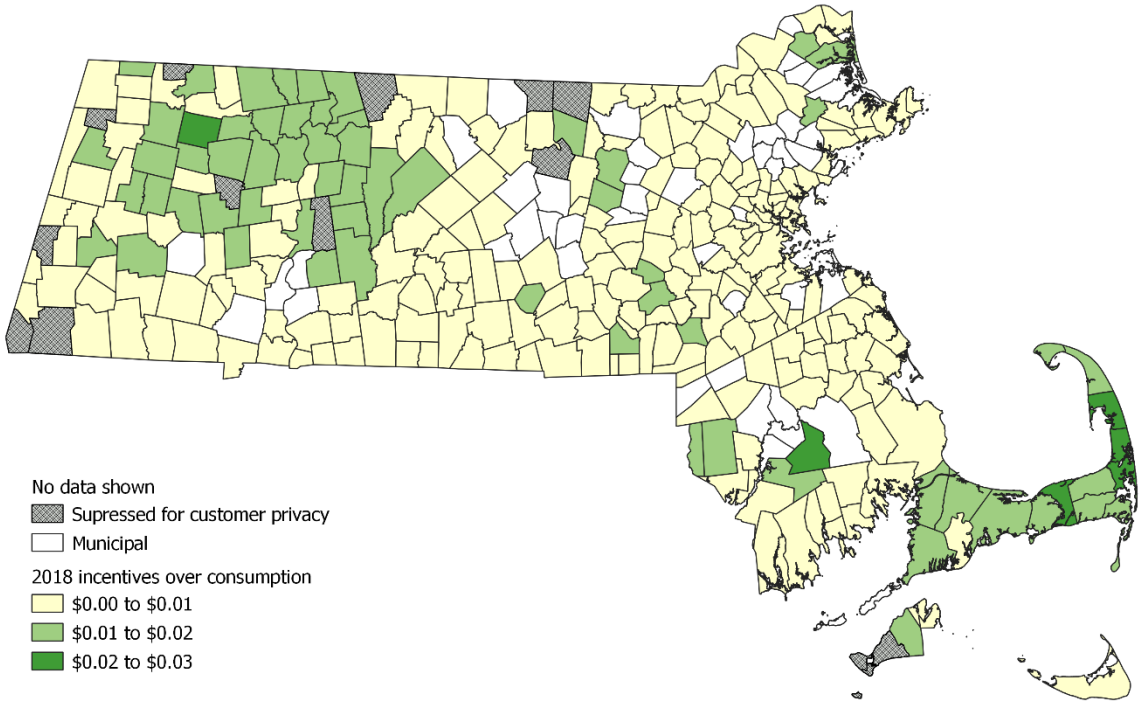
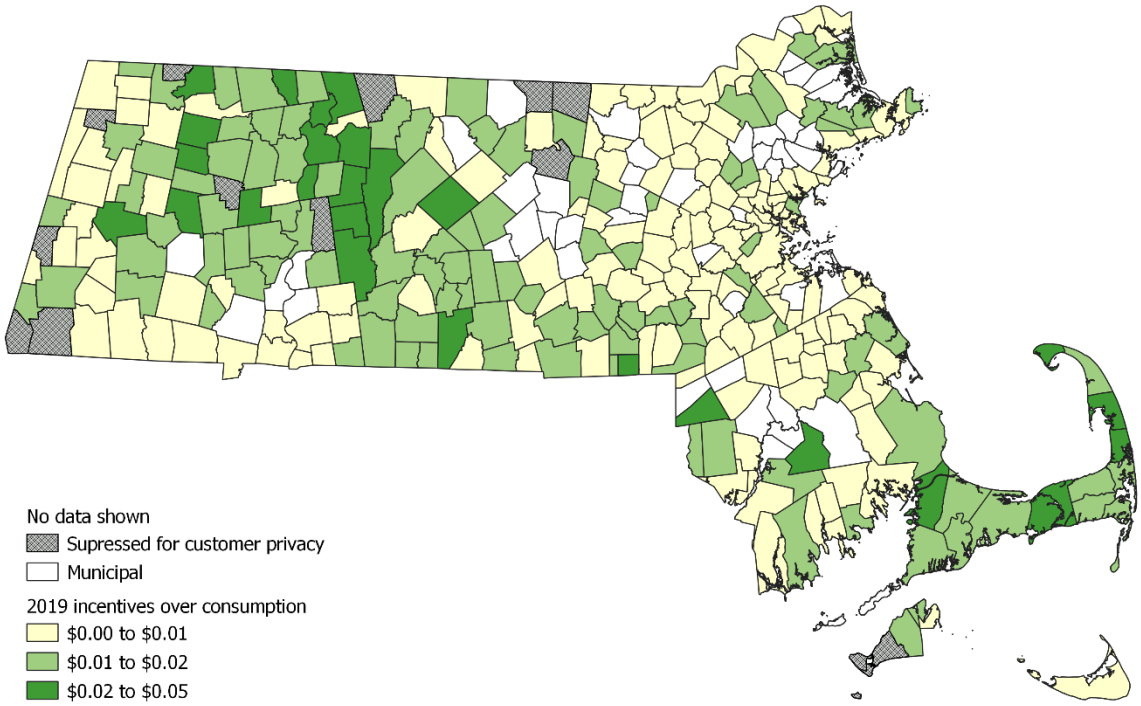


Figure 4-12. Map of incentive dollars per kWh consumption by town, 2019



4.3.2 Gas geographic analysis

Figure 4-13 shows incentive dollars spent cumulatively in 2013 through 2019 per kWh of consumption, aggregated to the town level. Lighter blue areas of the map show the areas with the lowest dollars spent per unit of energy usage, while darker blue areas show regions of higher spending. As with electric, there is no apparent straightforward relationship between consumption-weighted dollars spent and population size.

Figure 4-13. Map of incentive dollars per therm consumption by town, cumulative 2013-2019

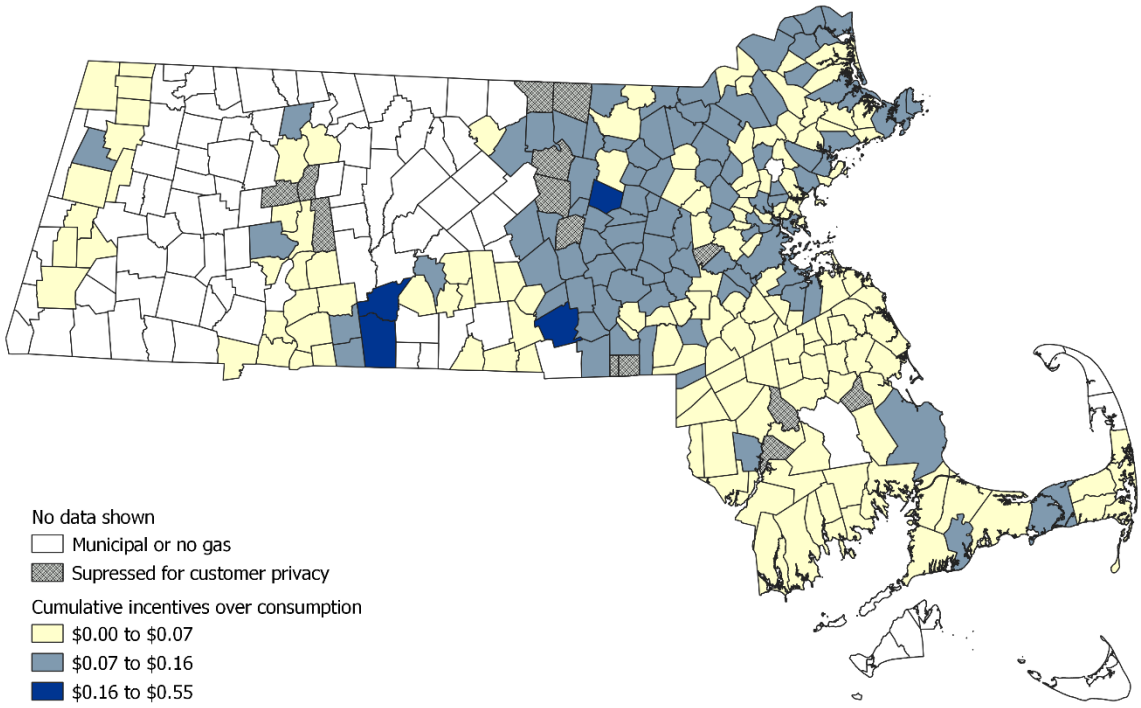


Figure 4-14 and Figure 4-15 show the incentive dollars spent per therm of consumption in 2018 and 2019, respectively. While there is some year-over-year variability, the overall pattern between years remains consistent with the longer-term view shown in Figure 4-13.

Figure 4-14. Map of incentive dollars per therm consumption by town, 2018

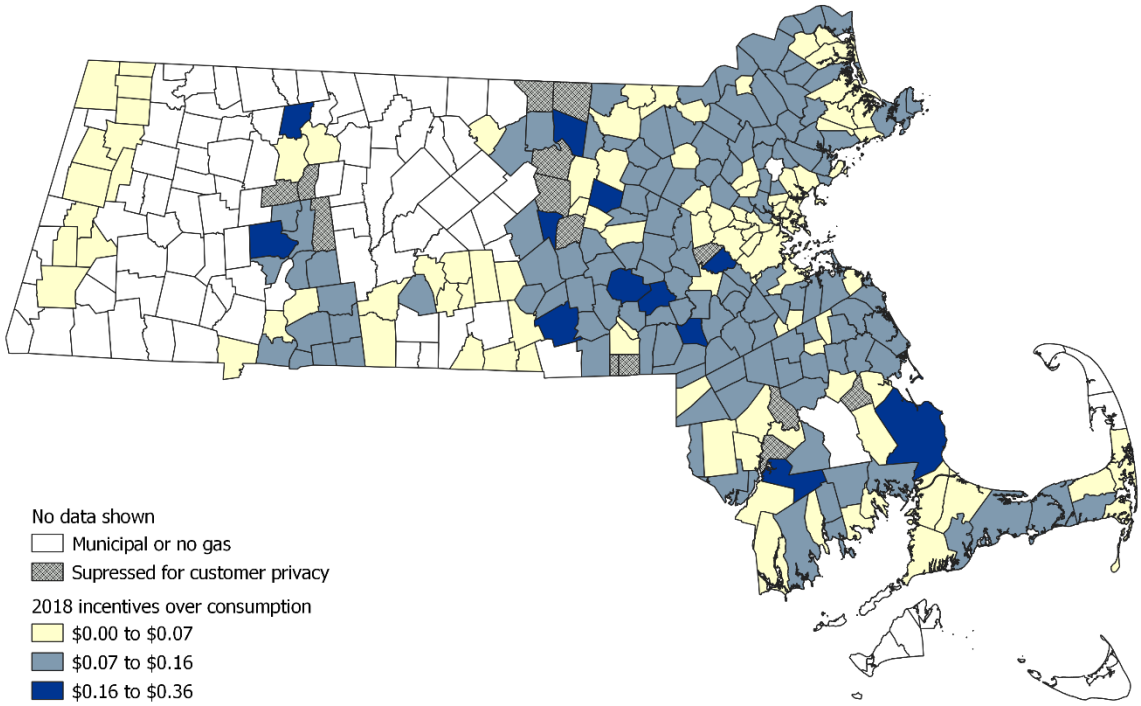
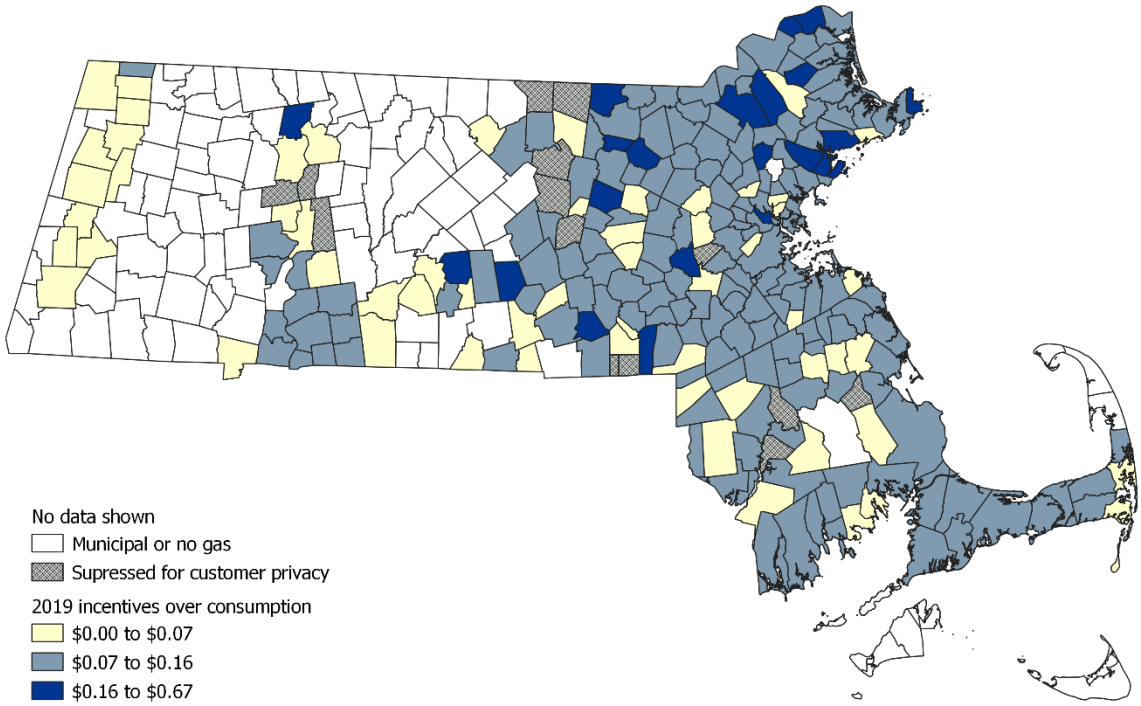


Figure 4-15. Map of incentive dollars per therm consumption by town, 2019



4.4 Renter Identification

The below tables show the overall renter percentages calculated by the improved renter algorithm described in Section 3.2. If any account or location was flagged by at least one algorithm, then we flagged the location or account as a renter. At the state-wide level, approximately 40% of accounts and 30% of locations are flagged as renters (Table 4-21).

DNV's method appears to be slightly under-counting renters at the statewide level. Locations are approximately the same as households in the ACS, and according to the 2019 ACS, 37.8% of households in MA are renters. This is partially caused by the conservative approach of our algorithm—if the algorithm does not have high confidence about renter status, the location/individual is not flagged as a likely renter. For example, the algorithm flags all condos as non-renter occupied, so without other factors indicating renter status, we will under-count rented condos. A renter will also not be flagged under the fuzzy name-matching algorithm unless there is a high probability that the names are dissimilar, so we will under-count renters whose names are like the name of the property owner. Finally, we will also under-count long-term renters with the account turnover flag. Therefore, where we identify a renter, we are confident of their renter status, but we have not captured the full renter population.

The difference between account and location level renter proportions reflects the fact that many renters reside in multi-unit buildings, which will be counted only once as a location, but multiple times by account.

Table 4-21. Statewide renter proportions

Level	Fuel Type	Renter
Account	Electric	38.9%
	Gas	39.0%
Location	Electric	33.7%
	Gas	31.0%

The renter proportions by PA are reported in Table 4-22 and Table 4-23. These are more difficult to compare to the ACS. However, they are generally close to the statewide values and the overall variance around those values is reasonable. There are areas where inconsistencies appear to persist. For example, Eversource provides most of the electricity in Boston, while National Grid provides most of the gas. Thus, we would expect National Grid gas percentages to be close to Eversource electric percentages, but they are not.

Table 4-22. Electric PA renter proportions

PA - Electric	Renter % by Location	Renter % by Account
National Grid	27.0%	31.7%
Eversource	43.0%	48.7%
Cape Light Compact	21.7%	26.9%
Unitil	29.0%	36.8%

Table 4-23. Gas PA renter proportions

PA - Gas	Renter % by Location	Renter % by Account
National Grid	21.1%	34.9%
Columbia	41.1%	44.2%
Eversource	37.4%	44.5%
Liberty	55.8%	56.6%
Berkshire	27.1%	25.9%
Unitil	24.4%	29.3%

Figure 4-16 shows a scatterplot of each block group based on the percentage of renters identified by DNV compared to the percentage of renters identified in the ACS. The overall correlation is 75%. As with the comparison of overall statewide percentages, the trend is for DNV to under-count renters in any given block group.

Figure 4-16. Block group renter percent; DNV method vs. ACS

In block groups where DNV's estimate is greater than the ACS (points above the main cluster of points), there are three explanations:

- We have flagged condos as renters where there is insufficient information in the PA and tax data to identify them as condos. (This is the most significant source of overcounting in our methodology). This contrasts with the under-counting of renter-occupied condos where we *can* identify a condo but assume that the residents are non-renters.
- The block group is within a municipal area that does not receive service from the PA, or there are very few PA accounts relative to the total number of locations in the block group. We are often able to identify PA-served areas on the edge of these block groups that are PA-served, and these often include high proportions of renters among the locations that the PAs serve. However, because DNV does not have any information about the other locations in the block group, those additional locations are not included in the denominators. Thus, the proportions are inflated relative to those in the ACS,

which does have full household counts in the denominators. For example, the highest outlier is a block group in a municipality where a large trailer park is the only PA-served portion in the residential data.

- We have identified student renters in single-family housing near colleges and universities that do not appear in the ACS data.

Where we are lower than the ACS, the most common cause is either unidentified housing authorities (which often have a single account holder and low turnover rates) that are not flagged by our algorithm and for which we do not have a tax use code, or other housing co-ops. For example, the lowest outlier is a senior rental community with long-term residents (low turnover), and little customer information in the utility data.

In general, we are confident that renters flagged by our algorithm are indeed renters, but we do not have high confidence that we have captured the full renter population.

4.5 Multifamily identification

Table 4-24 shows the results from this analysis, categories 5-9, 10-19, and 20+ all are within 1% of ACS. Buildings with 1, 2, and 3-4 units are all within 5% of ACS; a wider margin between ACS and DNV is expected because our analysis does not break out between attached and detached.

Table 4-24. Comparison of housing unit classifications between DNV's algorithm and the ACS estimates

Classification	Percentage of total housing units, DNV	Percentage of total housing units, ACS
1 Unit	52.1%	57.2%
2 Units	13.6%	9.3%
3-4 Units	13.0%	10.8%
5-9 Units	5.8%	5.8%
10-19 Units	3.8%	4.5%
20+ Units	11.6%	11.8%

We determined that while it was productive to compare the renter identification flag to the ACS at the block-group level, for the multifamily building size analysis, such comparisons do not yield useful results. This is because our analysis yields building size categories for locations for which there is only a single primary address listed in the PA data. So, for example, a single address classified as a 20+ unit building with 28 distinct units, for which we have only the primary address in the PA data, will sum to a single unit in aggregate (count of 20+ unit locations = 1) while the ACS would aggregate the same building to 28 units (count of 20+ unit housing units = 28). This is a limitation of the slight difference between the ACS definition of a housing unit and the street address, and the granularities at which these estimates are made.

An additional limitation to our approach is the use of electric-only data. This choice was made to limit the miscounting of master-metered multifamily buildings. However, it means that there are many locations for which we do not have address information in the electric data but do have address information from the tax data that are not included in the analysis.

We find that there is a correspondence between our building type classifications and the proportion of renters, as seen in Table 4-25. As expected, single-unit locations have the lowest renter proportion, followed by 2-unit. Noting that our approaches are conservative and count a location as renter-occupied only when there is high confidence, we see some correspondence with renter fractions from the ACS. However, these results suggest that our approach potentially overcounts single-unit locations, likely in master-metered multifamily buildings or where we have only the primary address with no sub-unit address information.

Table 4-25. Comparison of renter ratios by building size category between DNV and the ACS

Size category	Percentage renters, DNV	Percentage renters, ACS	Difference
1 unit	15%	10%	-5%
2 units	43%	60%	17%
3-4 units	68%	78%	10%
5-9 units	65%	84%	19%
10+	65%	82%	16%

We also performed extensive manual investigations, comparing building class assignments to street-view photographs and aerial imagery and classes to owner and customer names. We find that the building classifications through this algorithm are broadly accurate, though numerous edge cases are difficult to rigorously identify through any automated algorithm. For example, a single-unit condo association with numerous detached 1-unit buildings was classed by the tax data as a single location with 100+ units. Thus, we place lower confidence in these classifications than in the renter identification and note that this approach yields a "best guess" estimate only, albeit one that improves on using tax data alone.

4.6 Modeling

This section reports the results of the modeling analysis. It combines results from bivariate correlations and linear regression modeling. Bivariate correlations provide a basic idea of the high-level relationship between two variables. It answers simple questions such as, "Do renters participate less often than owners?" Modeling allows the analysis to consider nuances in the underlying data, most importantly controlling for the effects of demographic variables that tend to occur together. For example, households with low incomes also tend to rent. If both renters and low-income households have lower participation, modeling helps answer the question of whether it's the low income or the renting that is driving the effect on participation.

The section first presents block group-level findings separately for electric and gas. Within each fuel, the findings begin with a high-level, non-technical summary of the key takeaways of the analyses, followed by more in-depth, technical results. In the technical results sections, we show the bivariate correlations for each of the participation metrics and the demographic variables. Then we discuss the linear regression model that best predicts participation (the "best-fit" model). Finally, we present the results of a series of specific models run to better understand the effects of the interrelated demographic variables on participation.

The pattern described in the previous paragraph is then repeated for individual, account-level analyses.

4.6.1 Block group-level models

The first set of analyses examined relationships at the block group level of analysis. The questions answered by this level of analysis relate to relatively small geographic areas (Census block groups) as opposed to individual households. In other words, we explore how areas with particular characteristics participate more or less than areas with different characteristics. Unfortunately, the analysis does not answer whether individual buildings with those specific characteristics are more or less likely to participate. The block group-level results show the percentage of households in the block group that display a particular characteristic, such as the percentage of renters. The participation statistics represent similar, block group-level outcomes:

- *Block-group level location participation* is the percentage of locations in a block group served by the PAs that participated in a program at least once between 2013 and 2019. A location is approximately the same as a building, so this metric captures the breadth of participation. Depth of participation is less well captured: a multifamily building with

one unit that participates would count the same as a multifamily building with 50 units that participated. (Although such shallow participation in multifamily buildings is uncommon.)

- *Block group level consumption-weighted participation rate* is the same as location participation rate, except that it is weighted based on each participating building's total consumption. So, the numerator is the sum of consumption of all participating buildings and the denominator is the sum of consumption of all buildings, within a block group. This metric still represents the breadth of savings, but it better accounts for the relative size of each location than the unweighted location participation rate.
- *Block group-level population savings rate* is the sum of savings from all accounts within a block group from 2013 to 2019 divided by the sum of 2019 consumption from all accounts within a block group. This metric captures a combination of breadth and depth of savings because nonparticipants are represented as 0s in the numerator. The multifamily building with deeper participation would count more towards the numerator than the multifamily building with shallower participation.

Table 4-26. Key takeaways from the block group-level modeling - electric

Question	2019 Results	Analysis
How well do the demographic variables studied predict electric participation?	• The best combination of variables predicts approximately 23% of the variance of location participation (and lesser amounts of consumption-weighted location and savings rate). In other words, approximately three-fourths of the participation is driven by factors other than those considered in this study.	Best fit models
How does income relate to electric participation?	• Income is a strong predictor of electric participation. • Areas with higher concentrations of average or high income tend to have higher participation rates, across all three participation metrics. • Areas with higher concentrations of low- or moderate-income tend to have lower participation rates, across all three participation metrics. • The relationships for average or higher and moderate income hold even when controlling for overlapping demographics. • The relationship between low-income and population savings rate is driven by the overlap with renting.	All
How does ownership relate to electric participation?	• Ownership is the strongest predictor of electric participation. • Areas with higher concentrations of owners tend to have higher participation rates, across all three participation metrics. • Areas with higher concentrations of renters tend to have lower participation rates, across all three participation metrics. • These relationships hold even when controlling for overlapping demographics.	Correlations, A, C, G, H
How does housing type relate to electric participation?	• Housing type is an inconsistent predictor of electric participation. • Areas with higher concentrations of single-family housing tend to have higher participation rates, across all three participation metrics.	All

Question	2019 Results	Analysis
	• Areas with higher concentrations of multifamily or duplex, triplex, and fourplex housing tend to have lower participation rates, across all three participation metrics. • The relationship between population savings rates and multifamily housing appears to be partially driven by income and renter effects as well as the higher consumption inherent to larger buildings.	
How does limited English proficiency relate to electric participation?	• Areas with higher concentrations of limited English proficiency tend to have lower electric location participation rates, which is driven by overlaps with renting and low-income. • Areas with higher concentrations of limited English proficiency tend to have lower population savings rates, even after controlling for demographic overlaps. • Taken together, these findings suggest that even when households with limited English proficiency participate in programs, language barriers might be reducing their depth of savings.	Correlations, A, B, C, G
How do other demographic variables relate to electric participation?	• Areas with high concentrations of pre-1950 construction tend to have lower electric participation rates than areas with higher concentrations of post-1950 construction, across all three participation metrics. • The relationship between urban or rural status and electric participation rates is weak and inconsistent.	Correlations, C, G

Table 4-27. Key takeaways from the block group-level modeling - gas

Question	2019 Results	Analysis
How well do the demographic variables studied predict gas participation?	• The best combination of variables predicts approximately 24% of the variance of location participation (and lesser amounts of consumption-weighted location and savings rate). In other words, approximately three-fourths of the participation is driven by factors other than those considered in this study.	Best fit models
How does income relate to gas participation?	• Income is a strong predictor of gas participation. • Areas with higher concentrations of average or high income tend to have higher participation rates, across all three participation metrics. • Areas with higher concentrations of low- or moderate-income tend to have lower participation rates across all three participation metrics. • These relationships hold even when controlling for overlapping demographics	All

Question	2019 Results	Analysis
How does ownership relate to gas participation?	• Ownership is the strongest predictor of gas participation. • Areas with higher concentrations of owners tend to have higher participation rates, across all three participation metrics. • Areas with higher concentrations of renters tend to have lower participation rates, across all three participation metrics. • These relationships hold even when controlling for overlapping demographics.	Correlations, A, C, G, H
How does housing type relate to gas participation?	• Housing type is not a reliable independent predictor of gas participation. • Areas with higher concentrations of single-family housing tend to have higher participation rates, across all three participation metrics. • Areas with higher concentrations of multifamily or duplex, triplex, and fourplex housing tend to have lower participation rates, across all three participation metrics. However, these relationships are driven by overlaps with renter status.	All
How does limited English proficiency relate to gas participation?	• Areas with higher concentrations of limited English proficiency tend to have lower location participation rates, which is driven by overlaps with renting and low income. • Areas with higher concentrations of limited English proficiency tend to have lower population savings rates, even after controlling for demographic overlaps. • Taken together, these findings suggest that even when households with limited English proficiency participate in programs, language barriers might be reducing their depth of savings.	Correlations, A, B, C, G
How do other demographic variables relate to gas participation?	• Areas with high concentrations of pre-1950 construction tend to have lower participation rates than areas with higher concentrations of post-1950 construction, across all three participation metrics. • The relationship between urban or rural status and participation rates is weak and inconsistent.	Correlations, C, G

4.6.1.1 Electric location participation findings

Bivariate correlations

Figure 4-17 presents the between electric location-level participation and the 12 descriptive variables.

Key findings are:

- Areas with above-average incomes are more likely to participate and areas with below-average income are less likely to participate.
- Renter-occupied housing (and its inverse, Owner-occupied housing)[9] has the strongest correlation with electric location level participation. This is consistent with one of the main findings from the RNCSP—that the renter demographic is a key demographic for the PAs to attend to as they try to increase participation rates.
- Areas with a greater proportion of single-family are more likely to participate than areas with more duplexes-fourplexes or more multifamily housing.
- Areas with more homes built after 1950 are more likely to participate than areas with more homes built before 1950.
- More urbanization slightly increases participation rates.

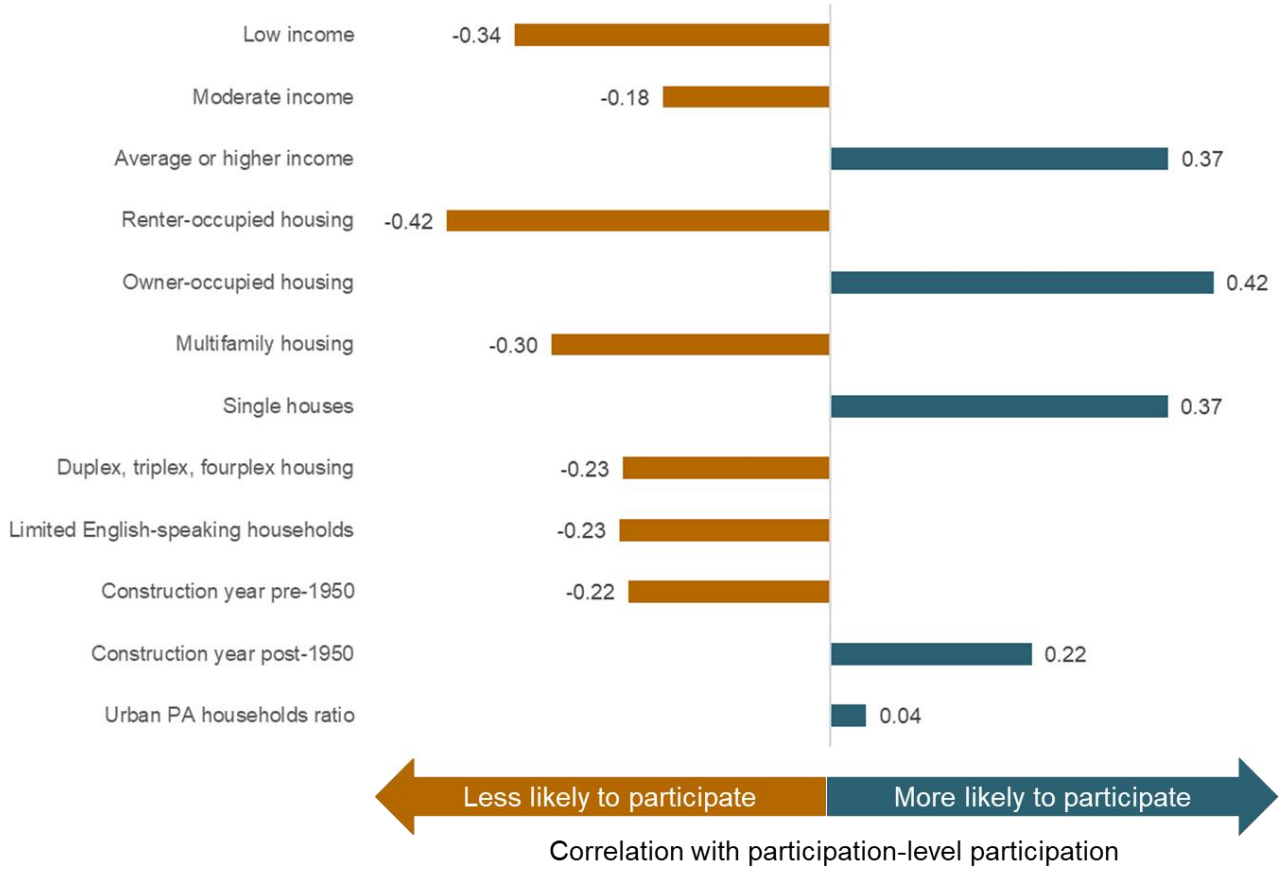
How to interpret a correlation

Bivariate correlations such as these show the relationship between participation and the demographic variables.

- A positive correlation indicates that as one variable increases, so does the other. In this case, the demographic variables are expressed in terms of the proportion of a block group that has that characteristic, so the interpretation of a positive correlation is that block groups with more of that demographic have greater participation.
- A negative correlation indicates the opposite pattern: as the proportion of a demographic variable increases, participation rate decreases. Or block groups with more of that demographic have lesser participation.
- The absolute value of the correlation coefficient indicates how strongly that pattern holds. An absolute value of 0 means that participation is unrelated to that demographic variable. An absolute value of 1 means the demographic variable perfectly predicts participation and is essentially indistinguishable. Absolute values less than 0.10 are generally considered very weak or negligible relationships, 0.30 to 0.70 are moderate, and absolute values above 0.70 are strong.

[9] The variables are set up such that each location is either renter occupied or owner occupied. Thus, each variable is the perfect inverse of the other. This is why the value of the correlation is the same for both variables, but with an opposite sign.

Figure 4-17. Correlation of electric location-level participation with ACS/US Census variables



Most of the demographic variables are intercorrelated. For example, renters are more likely to have low incomes than average or higher incomes (and vice versa). The full correlation matrix, including all demographic variables, is included in Appendix A.

Best-fit model

Figure 4-18 illustrates the progression of the R^2 as we built the best-fit model for electric location participation. The best-fit model for electric had 5 descriptive variables: renter-occupied housing, Urban PA households ratio, Average or higher income, Multifamily housing, and pre-1950 construction (Table 4-28). The total R^2 for this model was 0.226. This means the model was able to explain 22.6% of the variance of location-level participation, and the other 77.4% of the variance is unexplained by any of the demographic variables considered in this study.

Figure 4-18. R^2 progression of best-fit model – electric location participation

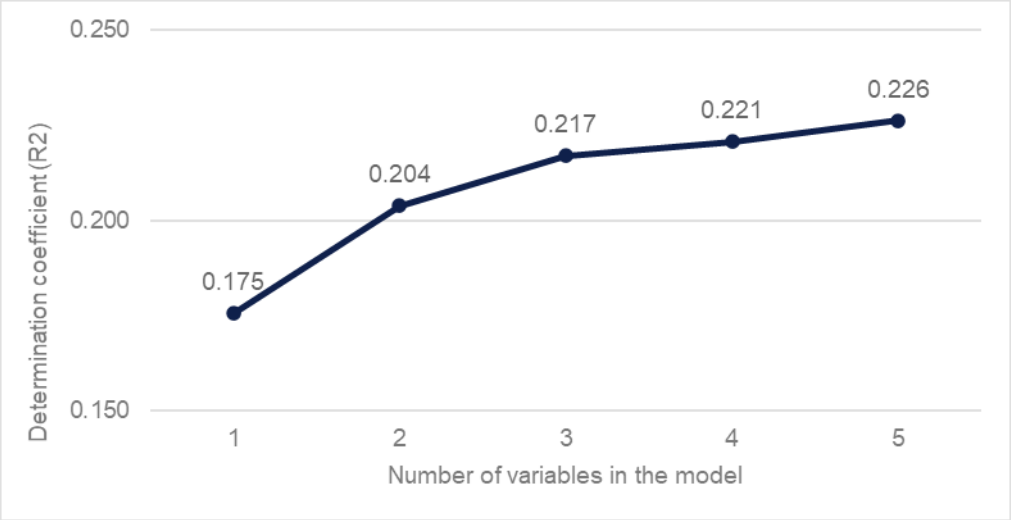


Table 4-28. Best-fit model – electric location participation

Descriptive Variable	Location-level participation	
	2019 model	
	Coefficients	p-value
Renter-occupied housing	-0.241	*0.000*
Urban PA households ratio	0.174	*0.000*
Average or higher income	0.169	*0.000*
Multifamily housing	-0.117	*0.000*
Construction year pre-1950	-0.092	*0.000*
	Total R^2 = 0.226	

When interpreting the models, it is important to remember that they show the effects of each descriptive variable while controlling for the overlaps with the other variables listed in the models. So, the effect of renter-occupied housing in this model is the effect of renting, independent of any increased (or decreased) likelihood of renters living in urban areas, having above-average incomes, living in multifamily housing, and living in older buildings.[10]

Like the conclusions reached in the RNCPS, the proportion of renter-occupied housing in a block group is the strongest single-variable predictor of electric location-level participation. Also consistent with the RNCPS, the relationship between renter-occupied housing and electric location-level participation is negative. This means that as the proportion of renter-occupied households in a block group increases, electric location-level participation tends to decrease. These models use standardized variables, so the coefficient of -0.241 means that for each 1 standard deviation increase in the proportion of

[10] Stated more precisely, because this is a block-group-level model, it is the effect of a high proportion of renters in a block group, independent of the increased/decreased likelihood that the block group is also more urban, has a higher proportion of above average income, has a greater concentration of multifamily housing, and has a greater concentration of pre-1950 construction.

renter-occupied housing, we predict a 0.241 standard deviation decrease in electric location participation rate. We cautiously interpret this to mean that renters are less likely to participate than owners.[11]

The positive coefficient for Urban PA household ratio indicates that more urbanized areas are more likely to participate than less urbanized areas, after controlling for any increases or decreases in renters, above-average income, multifamily housing, and pre-1950 construction are associated with urbanization. This is likely an effect of the increased administrative efficiency of targeting program activities at areas with greater population density.

Income also emerged as an important predictor of the electric location-level participation rate. In this case, average or higher income emerged as a positive predictor of participation. In other words, as the proportion of households with average or higher income increases, we also expect electric location-level participation to increase (by 0.174 standard deviations for each 1 standard deviation increase in average or higher income). With caution, we interpret this to mean that households with above-average incomes are more likely to participate than those with below-average incomes. This is a common phenomenon for energy efficiency programs across the country.

Multifamily housing had a negative coefficient. This means that block groups with higher concentrations of multifamily housing are less likely to participate, after controlling for any increases or decreases in renters, urbanization, above-average income, and pre-1950s construction. The specific models (Models A and B) run in the next section will investigate this interaction more closely.

We interpret the negative coefficient for pre-1950 construction to mean that older homes and buildings are less likely to participate than newer homes. DNV has found this effect in Massachusetts before. We attribute it to the higher prevalence of knob-and-tube wiring and/or asbestos in older construction that requires expensive remediation before energy efficiency improvements can be made.

Specific models

DNV ran additional, specific models based on those reported in the RNCPS. DNV used these models in the RNCPS to test specific hypotheses and data relationships. Repeating the models here allows us to test those same hypotheses and compare the results to the RNCPS.

These analyses focused on the effects of the variables that the PAs and EEAC agreed to pursue in the term sheet: customers with moderate income, renters, and households with limited English proficiency. Based on the results of the previous study, DNV also included other demographic variables such as housing type, low income, and date of construction. After an initial model (Model A) that tested the independent effects of these characteristics on participation, additional models tested nuances and interactions between the main term sheet characteristics and other demographic variables.

Model A

The purpose of Model A was to test the independent effects of the term sheet characteristics.

Table 4-29 presents the results of Model A for this study (2019 model) and the results reported in the RNCPS (2017 model) for comparison. Moderate income, renter-occupied housing, and multifamily housing all have negative relationships with electric location-level participation. This means we expect lower participation rates in block groups with more of these characteristics, after controlling for any overlaps in the likelihood of these characteristics appearing together in a block group. Despite the positive coefficient for limited English-speaking households, it should be interpreted as having no effect because it is not statistically significantly greater than 0.

[11] We approach this interpretation with caution because of the risk of "ecological fallacy". Relationships that hold at an aggregated level are not necessarily true of individual cases within the aggregate. We are comfortable making the cautious interpretation in this case because the RNCPS included individual-level models that were consistent with the block-group level models. Thus, we have some evidence that an ecological fallacy is unlikely in this case.

Thus, the 2019 Model A indicates that moderate income, renter-occupied housing, and multifamily housing are all independently associated with decreased participation rates. However, limited English proficiency is not related to an increase or decrease in participation rates.

Table 4-29. Model A – Term sheet and multifamily – Electric location participation

Descriptive Variable	Location-level participation			
	2019 model		2017 model	
	Coefficients	p-values	Coefficients	p-values
Intercept	----	----	0.389	<0.001
Moderate income	-0.103	0.000	-0.173	<0.001
Renter-occupied housing	-0.359	0.000	-0.178	<0.001
Multifamily housing	-0.070	0.000	0.081	<0.001
Limited English-speaking households	0.013	0.457	-0.027	n.s.

The effects in the 2019 model are very similar to those in the 2017 model. The differences are mostly attributable to the standardization process used this year compared to using non-standardized variables in 2017.

The one effect that changed relative to the previous study was the multifamily effect. This effect went from a (slight) positive predictor of participation in the previous study to a (slight) negative predictor in this study. This might also be attributable to an interaction between renters and multifamily and a nuanced difference in the interpretation of models that use standardized and non-standardized variables. In 2017, the multifamily effect reflects the effect of multifamily when there are no (0) renters in the block group. The 2017 model is saying that when there are no renters in a block group, an increase in the proportion of multifamily housing in the block group would tend to increase participation. Multifamily housing that is not rented probably is mostly condos.

Standardized variables are centered at the mean. This means that 0 for a standardized variable is the mean of that variable. So, the 2019 model, which uses standardized variables, indicates that when the proportion of renters is average (0), an increase in multifamily housing would tend to decrease participation. At average proportions of renters, multifamily housing can't be interpreted as condos.

Model B helps clarify the interaction with Renter-occupied housing and Multifamily housing.

Model B

Model B removes renter-occupied housing from Model A to test some of the interaction between renter-occupied housing and Multifamily housing (Table 4-30).

The best-fit model indicated that Renter-occupied housing is the strongest single variable predictor of participation rates. The effect of removing it from the model distributes that negative relationship among the remaining variables, commensurate with their overlap with Renter-occupied housing. All three of the remaining variables become much stronger, negative predictors of participation because they tend to overlap with renting and the redistribution of the Renter-occupied effect.

Taken together with the results of Model A, the Model B results indicate that Moderate income and Multifamily housing are related to lower participation rates in part because of the overlap with renting. However, they retain some significant predictive power independent of renting. Limited English proficiency has a negative relationship with participation because block groups with high concentrations of this characteristic also tend to have high concentrations of renters.

Table 4-30. Model B – Multifamily without Renter – Electric location participation

Descriptive Variable	Location-level participation			
	2019 model		2017 model	
	Coefficients	p-values	Coefficients	p-values
Intercept	----	----	0.363	<0.001
Moderate income	-0.151	*0.000*	-0.246	<0.001
Multifamily housing	-0.257	*0.000*	-0.032	<0.001
Limited English-speaking households	-0.130	*0.000*	-0.232	<0.001

The 2019 model is very similar to the 2017 model, and the differences can again be attributed to the standardization of the variables in the 2019 model. In particular, the interaction between renter-occupied housing and multifamily housing in the 2017 model becomes much clearer. The effects of multifamily housing became negative when renter-occupied housing was removed from the model. This is evidence of the interaction between renting and multifamily, and specifically demonstrates that in the 2017 analyses, multifamily housing negatively affected participation because of its overlap with renting. That effect is still present in the 2019 data along with a slight, independent negative effect of multifamily housing.

Model C

The final model tests the independent effects of all the descriptive variables by putting them all in the model at the same time. This model is practically the same as the best-fit model. It does not reveal any insights beyond those from the best-fit model.

Duplex, triplex, and fourplex housing appear for the first time in this model. The effect in this model is the independent effect of that housing type on participation, after removing the overlaps with the other demographics such as renting and income. In contrast, the bivariate correlation for duplex, triplex, fourplex housing was negative, when all the other overlapping demographic effects were added in. This pattern suggests that participation rates in duplex, triplex, fourplex housing are lower because of the overlaps with the other demographic variables.

Table 4-31. Model C – Effect of all descriptive variables – Electric location participation

Variable	Location-level participation	
	Coefficients	p-values
Intercept	----	----
Low income	-0.116	*0.000*
Moderate income	-0.095	*0.000*
Renter-occupied housing	-0.303	*0.000*
Multifamily housing	-0.079	*0.002*
Duplex, triplex, fourplex housing	0.079	*0.007*
Limited English-speaking households	-0.028	*0.113*
Construction year pre-1950	-0.115	*0.000*
Urban PA household ratio	0.174	*0.000*

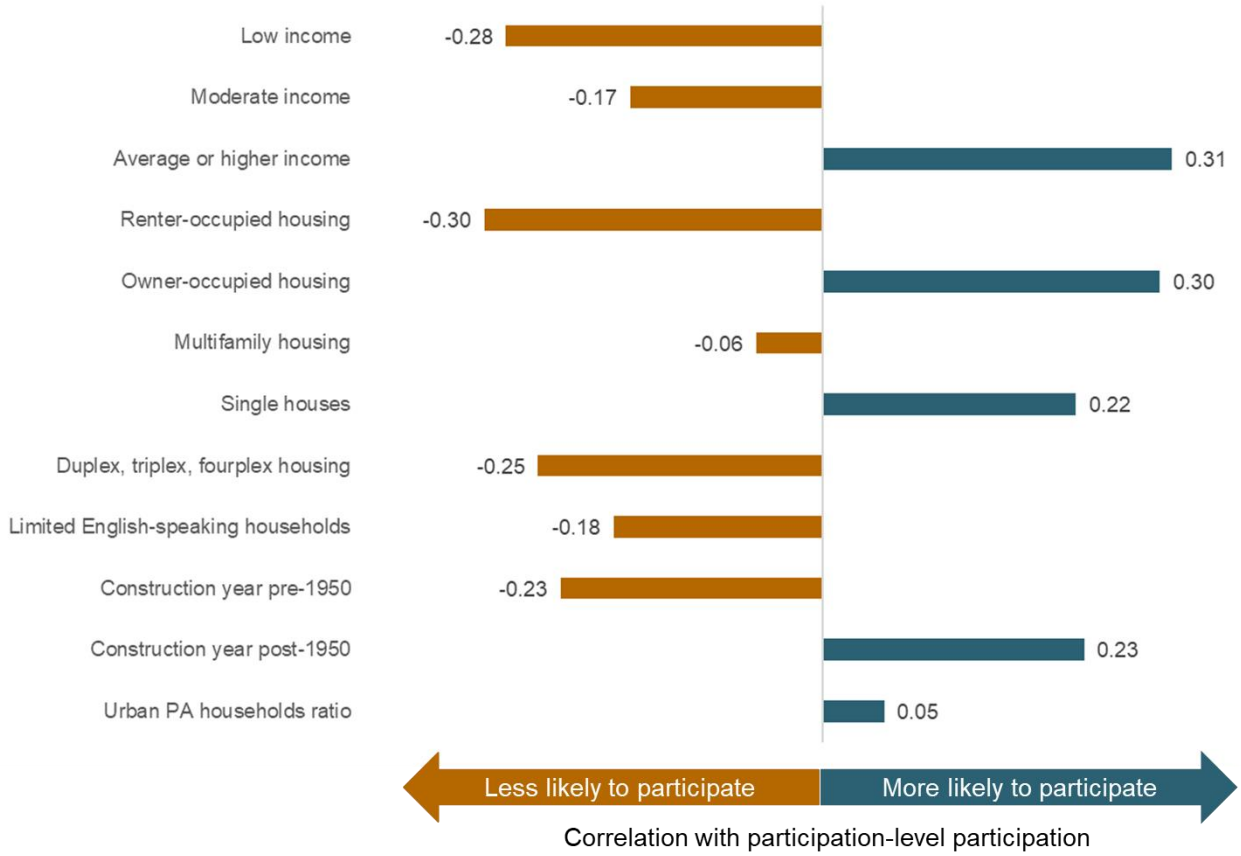
4.6.1.2 Electric consumption-weighted participation findings

The patterns of findings and implications for consumption-weighted location participation are the same as for unweighted location participation. The detailed results are included here for completeness and comparative purposes. Because these analyses add little to our understanding of participation patterns, our discussion is kept to a minimum.

Bivariate correlations

Figure 4-19 presents the correlations of electric consumption-weighted location participation with the demographic variables. Despite slight variations in magnitude from the unweighted electric location participation correlations, all patterns are the same.

Figure 4-19. Correlation of electric consumption-weighted participation with demographic variables



Best-fit model

The best-fit model for electric consumption weighted location participation (Table 4-32) is almost identical to the model for unweighted location participation. The descriptive variables load in a slightly different order, but they are all the same and have almost the same coefficients. The model explains less of the variance of consumption-weight participation.

The one effect that changes relative to the unweighted location participation best-fit model is that multifamily housing has a positive coefficient when weighted and negative co-efficiency when unweighted. This is partly an artifact of the weighting. Because multifamily buildings have such greater consumption than smaller buildings, the weighting greatly amplifies apparent participation rates. However, as will be confirmed in the population savings rate models, this result is also representative of the PAs achieving deeper savings in areas with greater concentrations of multifamily housing.

Table 4-32. Best-fit model – electric consumption-weighted location participation

Descriptive Variable	Consumption-weighted participation	
	2019 model	
	Coefficients	p-value
Average or higher income	0.169	*0.000*
Construction year pre-1950	-0.093	*0.000*
Urban PA households ratio	0.145	*0.000*
Renter-occupied housing	-0.300	*0.000*
Multifamily housing	0.173	*0.000*
	Total R^2 = 0.161	

Specific models

Model A

The positive relationship between multifamily housing and participation is stronger when participation is weighted by the total consumption within the building. This is attributable to the key limitation of consumption-weighted location participation rate– because multifamily buildings tend to have much greater consumption than single-family buildings, *any* participation in them gets amplified by that consumption, regardless of the depth of those savings.

Table 4-33. Model A – Term sheet and multifamily – Electric consumption-weighted location participation

Descriptive Variable	Consumption-weighted participation			
	2019 model		2017 model	
	Coefficients	p-values	Coefficients	p-values
Intercept	----	----	0.393	<0.001
Moderate income	-0.089	0.000	-0.164	<0.001
Renter-occupied housing	-0.426	0.000	-0.221	<0.001
Multifamily housing	0.221	0.000	0.280	<0.001
Limited English-speaking households	0.007	0.717	0.027	n.s.

Model B

The interaction between multifamily housing and renter-occupied housing manifests itself in Model B. The greater amount of consumption captured by treating multifamily buildings, the reduced likelihood that renters participate, and the overlap between renters and multifamily housing essentially cancel each other out to result in a non-significant effect for multifamily housing.

Table 4-34. Model B – Multifamily without Renter – Electric consumption-weighted location participation

Descriptive Variable	Consumption-weighted participation			
	2019 model		2017 model	
	Coefficients	p-values	Coefficients	p-values
Intercept	----	----	0.362	<0.001
Moderate income	-0.146	0.000	-0.255	<0.001
Multifamily housing	-0.001	0.973	0.139	<0.001
Limited English-speaking households	-0.164	0.000	-0.228	<0.001

Model C

Like the parallel model for unweighted location level electric participation, the consumption-weighted location level electric participation full model provides little additional information beyond the best-fit model.

Table 4-35. Model C – Effect of all descriptive variables – Electric consumption-weighted location participation

Variable	Consumption-weighted participation	
	Coefficients	p-values
Intercept	----	----
Low income	-0.120	0.000
Moderate income	-0.082	0.000
Renter-occupied housing	-0.406	0.000
Multifamily housing	0.244	0.000
Duplex, triplex, fourplex housing	0.138	0.000
Limited English-speaking households	-0.034	0.059
Construction year pre-1950	-0.136	0.000
Urban PA household ratio	0.142	0.000

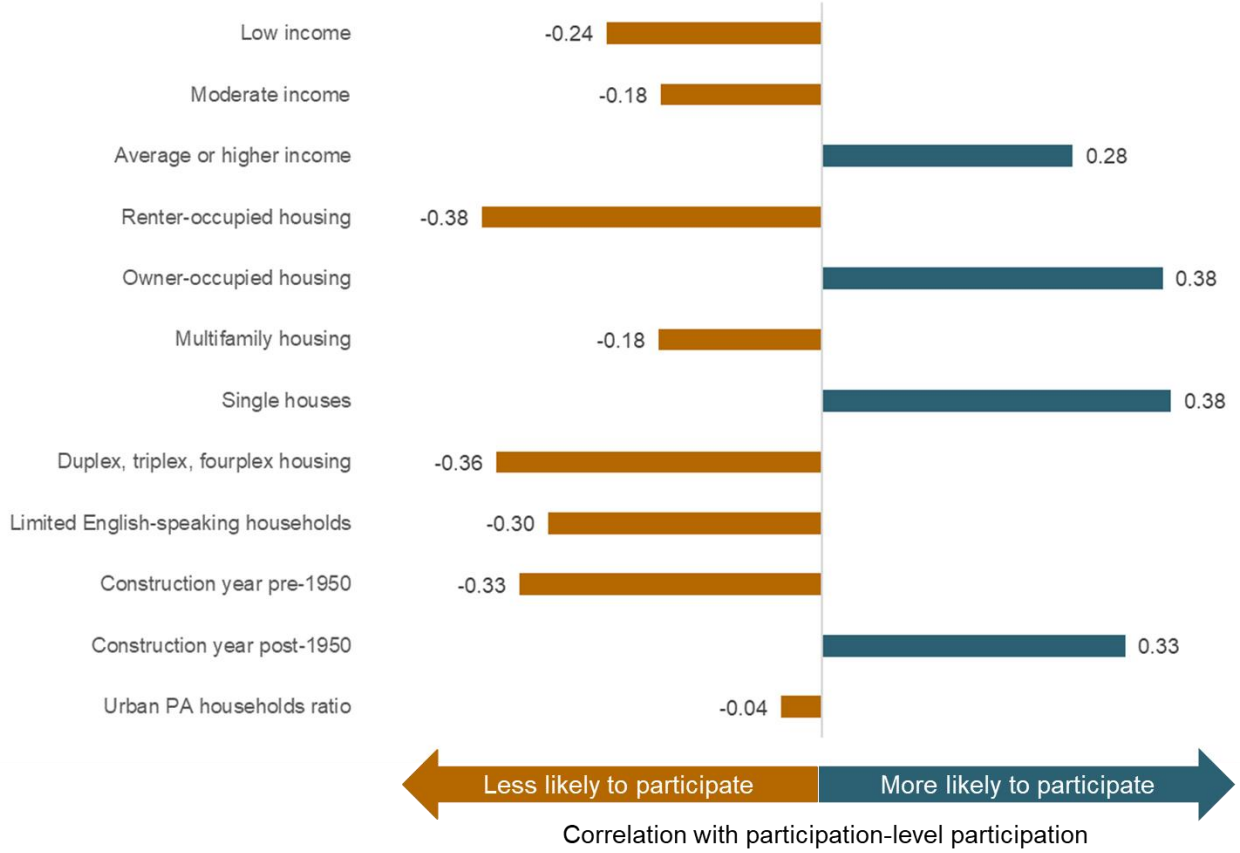
4.6.1.3 Electric population savings rate findings

The analyses on location-level participation provide insight into the breadth of participation. Analyses on population savings rate provide insight into the depth of savings. The analyses for population savings rate followed a similar procedure as for location level participation.

Bivariate correlations

Figure 4-20 presents the correlations of electric consumption-weighted location participation with the demographic variables. Despite slight variations in magnitude from the unweighted electric location participation correlations, the patterns are very similar. The only effect that is different is for the urban PA households ratio: more urbanized areas trend towards less depth of savings.

Figure 4-20. Correlation of electric population savings rate with demographic variables



Best-fit model

Figure 4-21 illustrates the progression of the R^2 as we built the best-fit model for the electric population savings rate. The best-fit model for electric had 5 descriptive variables: Single-family housing, Pre-1950 construction, Moderate income, Limited English-speaking households, and Urban PA households ratio (Table 4-36). The total R^2 for this model was 0.200. This means the model was able to explain 20% of the variance of population savings rate, and the other 80% of the variance is unexplained by any of the demographic variables considered in this study.

Figure 4-21. R^2 progression of best-fit model – electric population savings rate

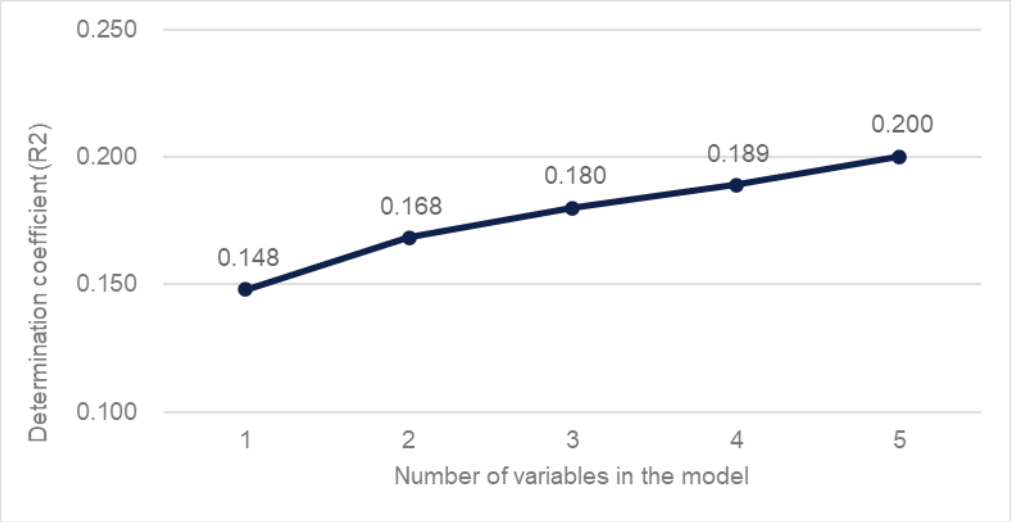


Table 4-36. Best-fit model – electric population savings rate

Descriptive Variable	Savings/consumption	
	2019 model	
	Coefficients	p-value
Single houses	0.229	0.000
Construction year pre-1950	-0.177	0.000
Moderate income	-0.098	0.000
Limited English-speaking households	-0.135	0.000
Urban PA households ratio	0.113	0.000
	Total R^2 = 0.200	

There are some differences between this best-fit model (which represents the depth of savings) and the location participation models (which indicate the breadth of savings.).

- Single-family housing is the most informative, single variable and has a high, positive coefficient. This indicates that savings are deepest for single-family homes. This effect is partially explained through the negative correlations between renting and location level participation and renting and single-family homes. Nonparticipating homes count as 0 savings in these analyses. Single-family homes are unlikely to be rented, and rented homes are unlikely to participate at all. Thus, if there are savings, it is more likely to be in single-family housing.
- The other implication of the Single-family housing effect is that when single-family homes participate, they tend to get relatively deeper savings than the other types of housing. Even though the PAs attempt to get all possible savings when they treat a multifamily building, there may either be less potential for these buildings, or there may be savings that the PAs are unsuccessful at achieving.
- Limited English fluency and Moderate income are both negatively related to the depth of savings, even after controlling for the effects of single-family housing (which indirectly partially controls for the overlaps of Limited English fluency and Moderate income with renting and multifamily housing). This implies that even when households from these demographics participate, the depth of savings achieved are lower than they are for fluent English speakers and the other income categories. The specific model results presented next help illuminate this finding.

- Like the location-level participation results, households from more urbanized areas tend to achieve greater savings.
- Also, like the location-level participation results, pre-1950 construction is associated with shallower savings. This is likely also related to the reduced likelihood of these buildings participating at all.

Specific models

As described for location-level participation rates, DNV ran additional, specific models to test specific hypotheses and data relationships among the term sheet characteristics and other descriptive variables.

Model G

The purpose of Model G was to test the independent effects of the term sheet characteristics. In the RNCPS (the 2017 model), we found that low income had a significant effect on the population savings rate, so we added it to this model, even though it is not officially one of the term sheet characteristics. We included it in the 2019 model to allow for direct comparison to the 2017 model. The non-significant effect for low income suggests that the other variables would continue to have similar effects as demonstrated in Model G if we removed Low income and reran the Model A on the population savings rate.

Table 4-37 shows the results of Model G for this study (2019 model) and the results reported in the RNCPS (2017 model) for comparison. Moderate income, renter-occupied housing, and limited English fluency all have negative relationships with electric population savings rates. This means we expect lower participation rates in block groups with more of these characteristics, after controlling for any overlaps in the likelihood of these characteristics appearing together in a block group.

The multifamily coefficient in this model is positive, but much less so than the coefficient in the parallel model on electric consumption-weighted location participation. Both effects are being influenced by the fact that consumption tends to be higher in multifamily buildings. In the case of savings rate, that greater consumption also contributes to greater savings.

The statistically significant and negative coefficients for limited English-speaking households and moderate-income are both notable here. Because Renter-occupied and Multifamily housing are also both in the model, these results help confirm the interpretation of the best-fit model that even when they participate, Moderate-income households and households with limited English proficiency are likely to have shallower savings than households in the other income categories or those with better English proficiency.

Table 4-37. Model G – Term sheet, low income, and multifamily – Electric population savings rate

Descriptive Variable	Savings/consumption			
	2019 model		2017 model	
	Coefficients	p-values	Coefficients	p-values
Intercept	----	----	0.062	<0.001
Low income	0.009	0.629	0.024	<0.001
Moderate income	-0.092	0.000	-0.051	<0.001
Renter-occupied housing	-0.346	0.000	-0.044	<0.001
Multifamily housing	0.081	0.000	0.048	<0.001
Limited English-speaking households	-0.119	0.000	0.054	<0.001

Model H

The purpose of Model H was to partially test interactions between limited English proficiency and the other term sheet characteristics. Table 4-38 shows the same effects as Model G after Limited English-speaking households are removed from the model. The coefficients in Model H are almost the same as those in Model G. Adding or removing Limited English

proficiency makes no noticeable difference on the effects of the other variables. This indicates the negative relationships due to income or place of residence with the depth of savings are not due to any overlap between those variables and limited English fluency.

Table 4-38. Model H – Term sheet, low income, and multifamily – Electric population savings rate

Descriptive Variable	Savings/consumption			
	2019 model		2017 model	
	Coefficients	p-values	Coefficients	p-values
Intercept	----	----	0.060	<0.001
Low income	0.009	0.634	0.029	<0.001
Moderate income	-0.093	0.000	-0.049	<0.001
Renter-occupied housing	-0.423	0.000	-0.036	<0.001
Multifamily housing	0.095	0.000	0.047	<0.001

Model C

The final model tests the independent effects of all the descriptive variables by putting them all in the model at the same time. This model is practically the same as the best-fit model. It does not reveal any insights beyond those from the best-fit model.

Table 4-39. Model C – Effect of all descriptive variables – Electric population savings rate

Variable	Savings/consumption	
	Coefficients	p-values
Intercept	----	----
Low income	-0.017	0.368
Moderate income	-0.091	0.000
Renter-occupied housing	-0.108	0.005
Multifamily housing	-0.068	0.008
Duplex, triplex, fourplex housing	-0.117	0.000
Limited English-speaking households	-0.119	0.000
Construction year pre-1950	-0.166	0.000
Urban PA household ratio	0.111	0.000

Additional discussion

In the bivariate correlations, low income was negatively correlated with the electric population savings rate. However, across all the models reported, low income did not have a statistically significant effect on the population savings rate. This pattern suggests that the negative effects of low income on the depth of savings are due to overlaps with the other demographic variables, such as renting.

In the RNCPS, DNV observed that the bivariate correlation between low income and electric population savings rate was positive. In that study, we attributed that finding to the success of the Low-income Multifamily (LIMF) programs at getting deep savings. When we removed LIMF participation from the RNCPS analyses, the low income–population savings rate correlation became negative rather than positive. We did not observe the same patterns in this study, so we cannot confirm the previous finding. However, the current pattern of findings regarding low income (persistent negative effects on the

breadth of participation combined with persistent non-effects on the depth of participation) could be due to a continuation of a weaker version of the same phenomenon.

4.6.1.4 Gas location-level participation findings

This section presents the results of electric location-level participation modeling. First, we present the bivariate correlations for each of the participation metrics and the demographic variables. Then we discuss the linear regression model that best predicts participation (the "best-fit" model). Finally, we present the results of a series of specific models run to better understand the interrelationships of the effects of the demographic variables on participation.

Bivariate correlations

Figure 4-22 presents the correlations between gas location-level participation and the 12 descriptive variables.

Key findings are:

- Areas with above-average incomes are more likely to participate and areas with below-average income are less likely to participate.
- Renter-occupied housing (and its inverse, Owner-occupied housing[12]) has the strongest correlation with gas location-level participation. This is consistent with one of the main findings from the RNCSP—that the renter demographic is a key demographic for the PAs to attend to as they try to increase participation rates.
- Areas with a greater proportion of single-family are more likely to participate than areas with more duplexes-fourplexes or more multifamily housing.
- Areas with more homes built after 1950 are more likely to participate than areas with more homes built before 1950.
- More urbanization slightly decreases participation rates.

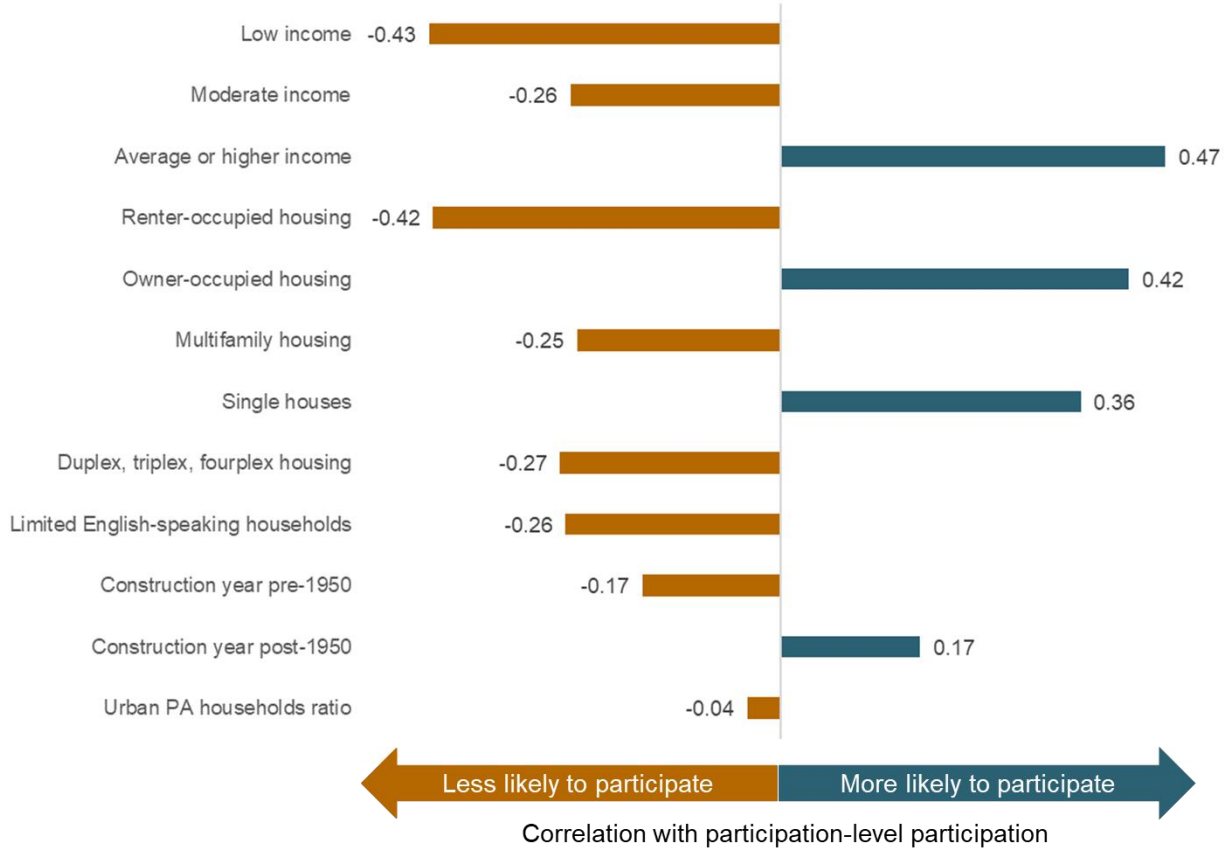
How to interpret a correlation

Bivariate correlations such as these show the relationship between participation and the demographic variables.

- A positive correlation indicates that as one variable increases, so does the other. In this case, the demographic variables are expressed in terms of the proportion of a block group that has that characteristic, so the interpretation of a positive correlation is that block groups with more of that demographic have greater participation.
- A negative correlation indicates the opposite pattern: as the proportion of a demographic variable increases, participation rate decreases. Or block groups with more of that demographic have lesser participation.
- The absolute value of the correlation coefficient indicates how strongly that pattern holds. An absolute value of 0 means that participation is unrelated to that demographic variable. An absolute value of 1 means the demographic variable perfectly predicts participation and is essentially indistinguishable. Absolute values less than 0.10 are generally considered very weak or negligible relationships, 0.30 to 0.70 are moderate, and absolute values above 0.70 are strong.

[12] The variables are set up such that each location is either renter occupied, or owner occupied. Thus, each variable is the perfect inverse of the other. This is why the value of the correlation is the same for both variables, but with an opposite sign.

Figure 4-22. Correlation of gas location-level participation with ACS/US Census variables



Most of the demographic variables are inter-correlated. For example, renters are more likely to have low incomes than average or higher incomes (and vice versa). The full correlation matrix, including all demographic variables, is included in Appendix A.

Best-fit model

Figure 4-23 illustrates the progression of the R^2 as we built the best-fit model for gas location participation. The best-fit model for electric had only two descriptive variables: Average or higher income, and Renter-occupied housing (Table 4-40). The total R^2 for this model was 0.241. This means the model was able to explain 24.1% of the variance of location-level participation, and the other 75.9% of the variance is unexplained by any of the demographic variables considered in this study.

The effects for average or higher income and renter-occupied housing are similar to their bivariate correlations. Both effects remain strong predictors of gas location participation, even after controlling for the negative correlation between income and renter status.

Figure 4-23. R^2 progression of best-fit model – gas location participation

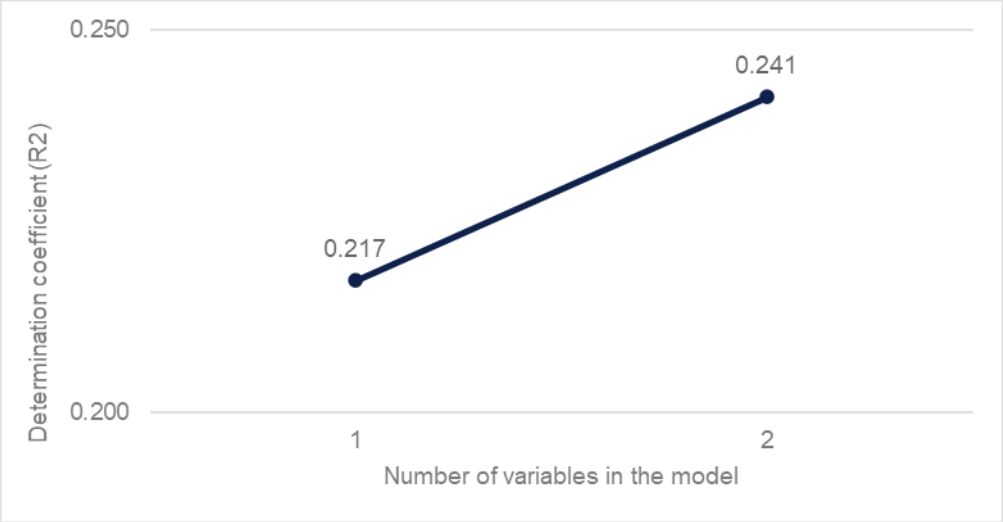


Table 4-40. Best-fit model – gas location participation

Descriptive Variable	Consumption-weighted participation	
	2019 model	
	Coefficients	p-value
Average or higher income	0.332	0.000
Renter-occupied housing	-0.205	0.000
	Total R^2 = 0.241	

Specific models

DNV ran additional, specific models based on those reported in the RNCPS. DNV used these models in the RNCPS to test specific hypotheses and data relationships. Repeating the models here allows us to test those same hypotheses and compare them to the results in the RNCPS.

These analyses focused on the effects of the variables that the PAs and EEAC determined characterize term sheet customers: moderate income, renters, multifamily, and households with limited English proficiency. After an initial model (Model A) that tested the independent effects of these characteristics on participation, additional models tested nuances and interactions between the main term sheet characteristics and other demographic variables.

Model A

The purpose of Model A was to test the independent effects of the term sheet characteristics.

Table 4-41 presents the results of Model A for this study (2019 model) and the results reported in the RNCPS (2017 model) for comparison. Moderate income and renter-occupied housing have independent negative relationships with gas location-level participation. This means we expect lower participation rates in block groups with more of these characteristics, after controlling for any overlaps in the likelihood of these characteristics appearing together in a block group with any of the other variables in the model. The effects of Multifamily housing and Limited English-speaking households were not statistically significantly different from zero. This suggests the effects of these variables are due to their overlap with renters. This interpretation is further illuminated in Model B.

Table 4-41. Model A – Term sheet and multifamily – Gas location participation

Descriptive Variable	Location-level participation			
	2019 model		2017 model	
	Coefficients	p-values	Coefficients	p-values
Intercept	----	----	0.339	<0.001
Moderate income	-0.158	0.000	-0.227	<0.001
Renter-occupied housing	-0.377	0.000	-0.198	<0.001
Multifamily housing	0.015	0.430	0.082	<0.001
Limited English-speaking households	-0.026	0.128	-0.105	<0.001

Model B

Model B removes renter-occupied housing from Model A to test some of the interaction between Renter-occupied housing and Multifamily housing (Table 4-42).

The best-fit model indicated that Renter-occupied housing is one of the strongest single variable predictors of participation rates. The effect of removing it from the model distributes that negative relationship among the remaining variables, commensurate with their overlap with Renter-occupied housing. All three of the remaining variables become much stronger, negative predictors of participation because they tend to overlap with renting and the redistribution of the Renter-occupied effect.

Taken together with the results of Model A, the Model B results indicate that Moderate income is related to lower participation rates in part because of the overlap with renting. However, it retains some significant predictive power independent of renting. Multifamily housing and Limited English proficiency have a negative relationship with participation because block groups with high concentrations of these characteristics also tend to have high concentrations of renters.

Table 4-42. Model B – Multifamily without Renter – Gas location participation

Descriptive Variable	Location-level participation			
	2019 model		2017 model	
	Coefficients	p-values	Coefficients	p-values
Intercept	----	----	0.313	<0.001
Moderate income	-0.211	0.000	-0.315	<0.001
Multifamily housing	-0.180	0.000	-0.046	<0.001
Limited English-speaking households	-0.177	0.000	-0.334	<0.001

The 2019 model is very similar to the 2017 model, and the differences can again be attributed to the standardization of the variables in the 2019 model. In particular, the interaction between Renter-occupied housing and Multifamily housing in the 2017 model becomes much clearer. The effects of Multifamily housing became negative when renter-occupied housing was removed from the model. This is evidence of the interaction between renting and multifamily, and specifically demonstrates that in the 2017 analyses, Multifamily housing negatively affected participation due to its overlap with renting. That effect is still present in the 2019 data along with a slight, independent negative effect of Multifamily housing.

Model C

The final model tests the independent effects of all the descriptive variables by putting them all in the model at the same time. This model is practically the same as the best-fit model. Urbanization has a slightly positive effect after controlling for

all the other variables, rather than a slightly negative effect in the bivariate correlations. This suggests that the negative effects of urbanization level are attributable to overlaps with the other demographic variables.

Table 4-43. Model C – Effect of all descriptive variables – Gas location participation

Variable	Location-level participation	
	Coefficients	p-values
Intercept	----	----
Low income	-0.255	0.000
Moderate income	-0.158	0.000
Renter-occupied housing	-0.241	0.000
Multifamily housing	0.017	0.489
Duplex, triplex, fourplex housing	0.020	0.503
Limited English-speaking households	-0.028	0.097
Construction year pre-1950	0.011	0.566
Urban PA household ratio	0.050	0.000

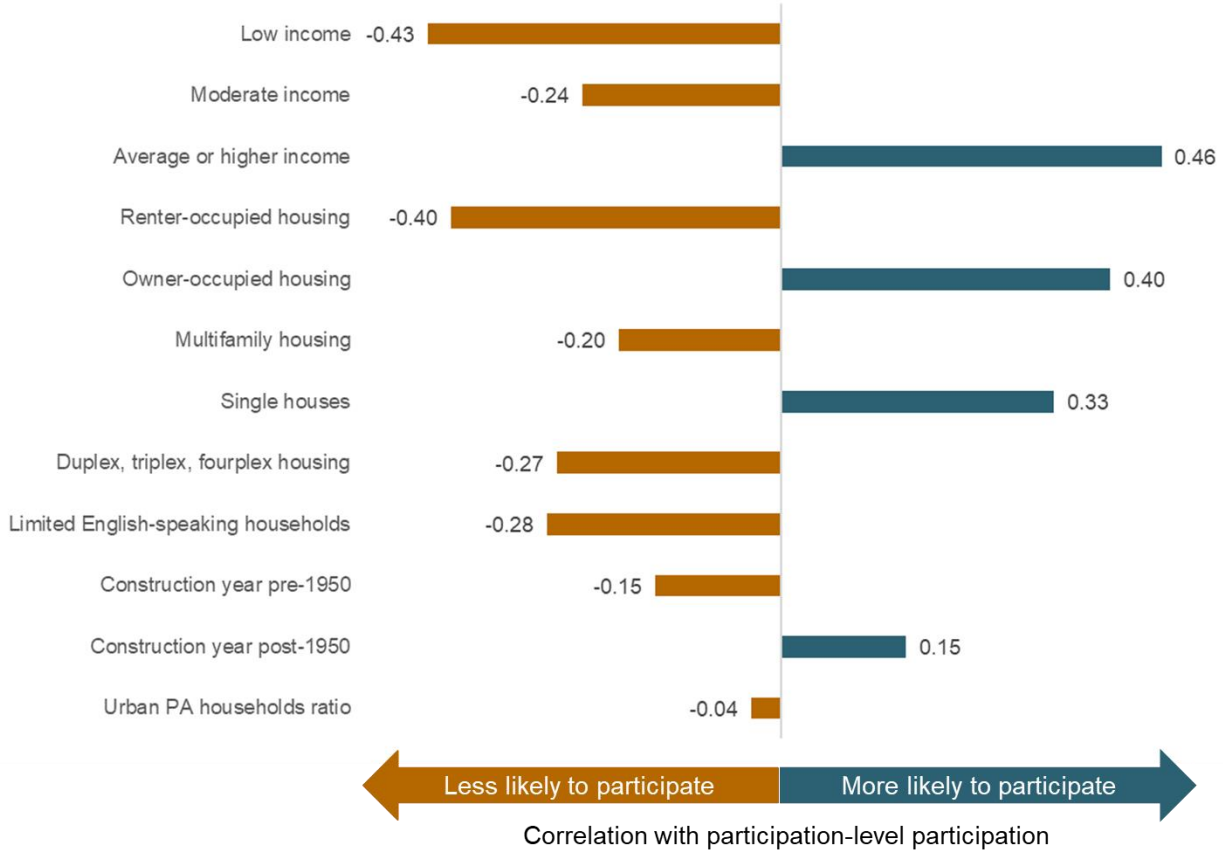
4.6.1.5 Gas consumption-weighted participation findings

The patterns of findings and implications for gas consumption-weighted location participation are the same as for unweighted gas location participation. The detailed results are included here for completeness and comparative purposes. Because these analyses do little to add to our understanding of participation patterns, our discussion is kept to a minimum.

Bivariate correlations

Figure 4-24 presents the correlations of electric consumption-weighted location participation with the demographic variables. Despite slight variations in magnitude from the unweighted electric location participation correlations, all patterns are the same.

Figure 4-24. Correlation of gas consumption-weighted participation with demographic variables



Best-fit model

The best-fit model for gas consumption-weighted location participation (Table 4-44) is very similar to the model for unweighted gas location participation. In this case, the limited English-speaking household variable contributes explanatory power above and beyond Average or higher income and Renter-occupied housing. Despite the additional variable, the model explains slightly less of the variance of consumption-weighted participation.

Table 4-44. Best-fit model–gas consumption-weighted location participation

Descriptive Variable	Location-level participation	
	2019 model	
	Coefficients	p-value
Average or higher income	0.347	0.000
Renter-occupied housing	-0.133	0.000
Limited English-speaking households	-0.070	0.000

Total R^2 = 0.233

Specific models

Model A

As observed in the unweighted gas location participation Model A, the effects of Multifamily housing and Limited English-speaking households are flipped or muted relative to the bivariate correlations when Renter-occupied housing is in the model. This pattern suggests that the negative effects of Multifamily housing and Limited English-speaking households are attributable to the overlaps with renting.

Table 4-45. Model A – Term sheet and multifamily – gas consumption-weighted location participation

Descriptive Variable	Consumption-weighted participation			
	2019 model		2017 model	
	Coefficients	p-values	Coefficients	p-values
Intercept	----	----	0.358	<0.001
Moderate income	-0.142	0.000	-0.217	<0.001
Renter-occupied housing	-0.369	0.000	-0.215	<0.001
Multifamily housing	0.071	0.000	0.163	<0.001
Limited English-speaking households	-0.070	0.000	-0.197	<0.001

Model B

The interaction between Multifamily housing and Renter-occupied housing is more apparent in Model B. When renter-occupied housing is not in the model, Multifamily housing continues to have a negative relationship with gas consumption-weighted location participation.

Table 4-46. Model B – Multifamily without Renter – gas consumption-weighted location participation

Descriptive Variable	Consumption-weighted participation			
	2019 model		2017 model	
	Coefficients	p-values	Coefficients	p-values
Intercept	----	----	0.331	<0.001
Moderate income	-0.193	0.000	-0.317	<0.001
Multifamily housing	-0.120	0.000	0.024	<0.01
Limited English-speaking households	-0.218	*0.000*	-0.448	<0.001

Model C

Table 4-47 shows the model outcomes when all demographic variables are included in the model. These results do not provide any additional insights beyond the previously reviewed models.

Table 4-47. Model C – Effect of all descriptive variables – gas consumption-weighted location participation

Variable	Consumption-weighted participation	
	Coefficients	p-values
Intercept	----	----
Low income	-0.276	0.000
Moderate income	-0.142	0.000
Renter-occupied housing	-0.238	0.000
Multifamily housing	0.083	0.001
Duplex, triplex, fourplex housing	0.019	0.528
Limited English-speaking households	-0.072	0.000
Construction year pre-1950	0.039	0.041
Urban PA household ratio	0.051	0.000

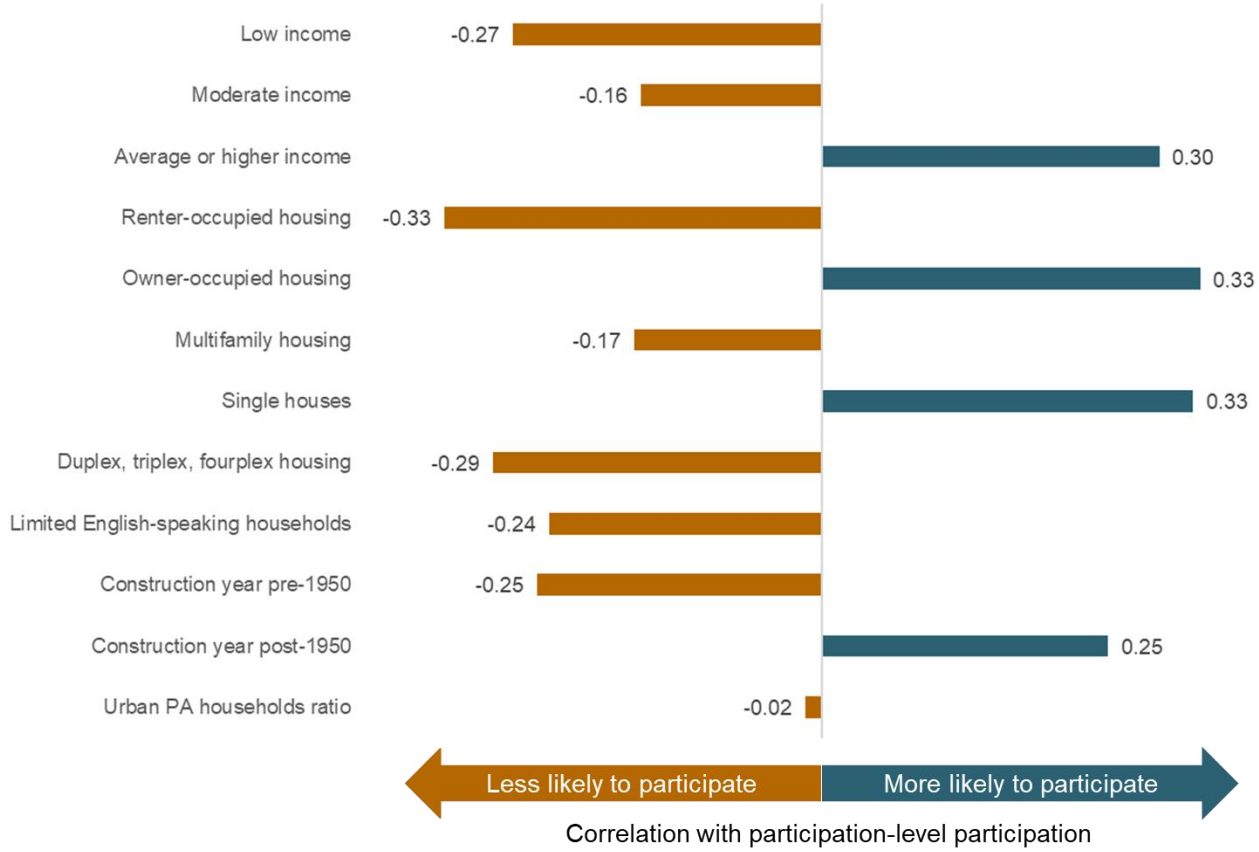
4.6.1.6 Gas population savings rate findings

This section presents the results of savings/consumption variable modeling. First, we discuss the correlation of individual variables, followed by a presentation and discussion of the model built with the use of ACS and US Census variables. Finally, we recreate models from previous versions of this document and compare the results year to year.

Bivariate correlations

Figure 4-25 presents the correlations of electric consumption-weighted location participation with the demographic variables. Despite slight variations in magnitude from the unweighted electric location participation correlations, all patterns are the same.

Figure 4-25. Correlation of gas population savings rate with demographic variables



Best-fit model

Figure 4-26 illustrates the progression of the R^2 as we built the best-fit model for the gas population savings rate. The best-fit model for gas had 5 descriptive variables: Renter-occupied housing, Pre-1950 construction, Average or higher income, Urban PA households ratio, and Limited English-speaking households (Table 4-48). The total R^2 for this model was 0.144. This means the model was able to explain 14.4% of the variance of population savings rate, and the other 85.6% of the variance is unexplained by any of the demographic variables considered in this study.

Figure 4-26. R^2 progression of best-fit model – gas population savings rate

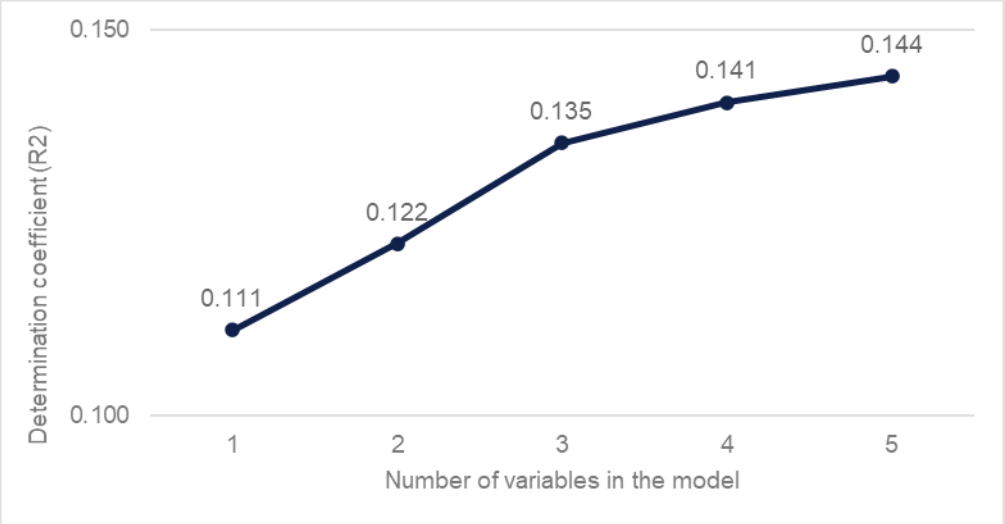


Table 4-48. Best-fit model – gas population savings rate

Descriptive Variable	Savings/consumption	
	2019 model	
	Coefficients	p-value
Renter-occupied housing	-0.151	0.000
Construction year pre-1950	-0.137	0.000
Average or higher income	0.146	0.000
Urban PA households ratio	0.081	0.000
Limited English-speaking households	-0.070	0.000
	Total R^2 = 0.144	

Renter-occupied housing is the most informative, single variable and has a high, negative coefficient. This effect is partially explained through the negative correlations between renting and gas location-level participation because nonparticipating homes count as 0 savings in these analyses.

We interpret the negative coefficient for pre-1950 construction indicates that older homes and buildings are less likely to participate than newer homes. DNV has found this effect in Massachusetts before. We attribute it to the increased frequency of expensive remediation needed in older construction before energy efficiency improvements can be made.

Average or higher income has a positive relationship with gas population savings rate. This effect is independent of the other variables in the model, which partially account for the reduced likelihood of any participation. Thus, this effect suggests that when they do participate, homes with average or higher incomes achieve a greater depth of savings than participating homes with below-average incomes.

Limited English fluency is negatively related to the depth of savings, even after controlling for the effects of the other variables (which indirectly partially controls for the overlaps of Limited English fluency and Moderate income with renting and multifamily housing). This implies that even when households from these demographics participate, the depth of savings achieved are lower than they are for fluent English speakers. The specific model results presented next help illuminate this finding.

Like the electric results, households from more urbanized areas tend to achieve greater savings.

Key insights provided by this model are listed below:

- Renter-occupied housing's coefficient suggests that renters generate lower savings compared to consumption than non-renters. This is consistent with all previously presented analyses.
- As the 0-order correlation analysis suggested, the fact that a building was built in a year pre-1950 implies reduced savings generation. Because non-participants are counted as 0s in this analysis, this effect could be caused by a lower participation rate among older homes (often caused by the need for extra, more expensive remediations). From a pure depth of savings perspective, it could also be due to a tendency for older homes to be smaller.
- The income effect might also be attributable to the decreased participation rates among households with low or moderate incomes. Homes among the lower-income groups might also tend to be smaller.
- Urban areas generate higher savings, whereas English limited proficiency is a factor that reduces the tendency of households to save energy.

Specific models

As described for location-level participation rates, DNV ran additional, specific models to test specific hypotheses and data relationships among the term sheet characteristics and other descriptive variables.

Model G

The purpose of Model G was to test the independent effects of the term sheet characteristics. In the RNCPS (the 2017 model), we found that low income had a significant effect on the population savings rate, so we added it to this model, even though it is not officially one of the term sheet characteristics. We included it in the 2019 model to allow for direct comparison to the 2017 model.

Table 4-49 shows the results of Model G for this study (2019 model) and the results reported in the RNCPS (2017 model) for comparison. All the tested variables have independent, negative relationships with gas population savings rates. This means we expect lower participation rates in block groups with more of these characteristics, after controlling for any overlaps in the likelihood of these characteristics appearing together in a block group.

The multifamily coefficient in this model is positive, while the bivariate correlation is negative. This pattern indicates that the negative effect of multifamily housing on the gas participation savings rate is attributable to the overlap with the other demographic variables in this model.

The statistically significant and negative coefficient for Limited English-speaking households is notable here. Because Renter-occupied and Multifamily housing are also both in the model, these results help confirm the interpretation of the best-fit model that even when they participate, households with limited English proficiency are likely to have shallower savings than households in the other income categories or those with more English proficiency.

Table 4-49. Model G – Term sheet, low income, and multifamily – gas population savings rate

Descriptive Variable	Savings/consumption			
	2019 model		2017 model	
	Coefficients	p-values	Coefficients	p-values
Intercept	----	----	0.057	<0.001
Low income	-0.092	0.000	-0.003	n.s.
Moderate income	-0.076	0.000	-0.044	<0.001
Renter-occupied housing	-0.253	0.000	-0.036	<0.001
Multifamily housing	0.055	0.004	0.046	<0.001
Limited English-speaking households	-0.065	0.000	0.030	<0.01

Model H

The purpose of Model H was to partially test interactions between limited English proficiency and the other term sheet characteristics. Table 4-50 shows the same effects as Model G after Limited English-speaking households are removed from the model. The coefficients in Model H are almost the same as those in Model G. Adding or removing Limited English proficiency makes no noticeable difference on the effects of the other variables. This indicates the negative relationships due to income or place of residence with the depth of savings are not due to any overlap between those variables and limited English fluency.

Table 4-50. Model H – Term sheet, low income, and multifamily – gas population savings rate

Descriptive Variable	Savings/consumption			
	2019 model		2017 model	
	Coefficients	p-values	Coefficients	p-values
Intercept	----	----	0.056	<0.001
Low income	-0.094	0.000	0.000	n.s.
Moderate income	-0.077	0.000	-0.043	<0.001
Renter-occupied housing	-0.294	0.000	-0.031	<0.001
Multifamily housing	0.065	0.001	0.045	<0.001

Model C

Table 4-51 shows the model outcomes when all demographic variables are included in the model. These results do not provide any additional insights beyond the previously reviewed models.

Table 4-51. Model C – Effect of all descriptive variables – gas population savings rate

Variable	Savings/consumption	
	Coefficients	p-values
Intercept	----	----
Low income	-0.115	0.000
Moderate income	-0.079	0.000
Renter-occupied housing	-0.090	0.025
Multifamily housing	-0.041	0.117
Duplex, triplex, fourplex housing	-0.075	0.015
Limited English-speaking households	-0.063	0.000
Construction year pre-1950	-0.114	0.000
Urban PA household ratio	0.082	0.000

4.6.2 Individual-level models

The advantage of individual-level modeling is that the results pertain to the characteristics of specific buildings rather than geographic areas that contain many buildings. Thus, the individual models can answer questions such as whether buildings with specific characteristics are more or less likely to participate in energy efficiency programs. The limitation of these analyses is that the predictor variables are much more limited than in the case of the block group-level analyses.

The dependent variables for the individual-level models are as follows:

- Location participation is 1 if any account associated with that location (building) participated at least once between 2013 and 2019.
- The location savings rate is the sum of savings from 2013 to 2019 divided by the average consumption from 2013 to 2019 at a location, for locations that participated at least once between 2013 and 2019.

The predictor variables represented 24 categories of buildings based on 3 characteristics:

- Pre-1950 construction or post-1950 construction
- Rented or owned
- Six size categories: 1 unit, 2 units, 3-4 units, 5-9 units, 10-19 units, and 20+ units

The key takeaways from the individual-level modeling are listed in Table 4-52 and Table 4-53.

Table 4-52. Individual-level modeling key takeaways – electric

Question	2019 Results	Analysis
How do participation rates vary across building sizes?	• Single-family buildings are most likely to participate. This finding is consistent with the results of the block group-level models. • 5-9 unit buildings are least likely to participate.	Participation ANOVA
How do participation rates vary across ownership?	• Owners are more likely to participate than renters. • This finding is consistent with the block group-level models.	
How do participation rates vary across building age?	• Newer buildings are more likely to participate than older buildings. • This finding is consistent with the block group-level models	
How do the building characteristics interact to predict participation?	• Within single-family and duplexes, ownership and building age affect participation rates. • Within triplexes and fourplexes, newer buildings are more likely to participate than older buildings, but ownership does not predict participation rates. • Within multifamily buildings (5+ units), both ownership and building age matter. Older, rented buildings are more likely to participate than newer, owned buildings.	
When a building participates, what is the best predictor of the amount of savings?	• Building size is the most consistent predictor of the depth of savings with a significant result in 95% of the Monte Carlo runs. Participating single-family buildings have the greatest depth of savings, and 5-9 unit buildings have the least. Triplex-fourplex buildings also have a low depth of savings. • The interaction between building size and ownership is also a relatively consistent predictor of savings with significant results in 18% of the Monte Carlo runs. The nature of this relationship is that there is little difference in depth of savings for renters versus owners in single-family buildings, and a substantial difference between	Savings ANOVA

Question	2019 Results	Analysis
	renters and owners for 10+ unit buildings. Renters in the 10+ unit buildings achieve deeper savings than owners.	

Table 4-53. Individual-level modeling key takeaways – gas

Question	2019 Results	Analysis
How do participation rates vary across building sizes?	• Single-family buildings are the most likely to participate. • 5-9 unit buildings are the least likely to participate.	Participation ANOVA
How does participation rate vary across ownership?	• Owners are more likely to participate than renters. • This is consistent with the block group-level analyses.	
How does participation rate vary across building age?	• Newer buildings are more likely to participate than older buildings. • This is consistent with the block group-level results.	
How do the building characteristics interact to predict participation?	• Within single-family buildings, duplexes, and triplexes/fourplexes, both ownership and building age have small effects. Owners and newer buildings are more likely to participate. • In 5-9 unit buildings, both ownership and building age make substantial differences. Owners and newer buildings are more likely to participate. • For 10-19 unit buildings, owners of newer buildings (likely condos) are more likely to, and the other 4 combinations are less likely to participate. • Within 20+ unit buildings, renters of newer buildings are the least likely to participate, and the other 3 combinations are the most likely to participate across all 24 categories.	
When a building participates, what is the best predictor of the amount of savings?	• Ownership is the best predictor of the depth of savings for participating gas buildings, with a significant effect in 65% of the Monte Carlo runs. Renters have deeper savings (ratio=14.0) than owners (ratio=11.2). This effect appears to be driven by multifamily (5+ unit buildings). In multifamily buildings, renters save much more than owners; in single-family through fourplex buildings, there is little difference in the depth of savings for renters versus owners.	Savings ANOVA

4.6.2.1 Electric participation rate

The participation analysis estimated the probability of an individual location participating in an energy efficiency program at least once between 2013 and 2019 for each of the 24 location categories (pre/post 195 construction * renter/owner * six building size categories).[13] The probability did not exceed 50% for any category (Table 4-54).

Table 4-54. Location-level probability of participation - electric

	Single family	Duplex	Triplex-fourplex	5-9 units	10-19 units	20+ units	Total
Owners							
Built post 1950	33%	25%	28%	31%	41%	40%	**33%**
Built pre 1950	33%	24%	20%	21%	21%	26%	**31%**
Renters							
Built post 1950	32%	28%	27%	16%	23%	21%	**29%**
Built pre 1950	30%	27%	21%	12%	17%	15%	**25%**
Total	**33%**	**26%**	**24%**	**17%**	**26%**	**22%**	

The main effects are very consistent with the block group-level results:

- Across the building size category, single-family homes are the most likely to participate (33% probability). Five to 9-unit buildings are least likely to participate (17% probability). The other building size categories have participation probabilities ranging from 22% for the 20+ unit buildings to 26% for duplexes and 10-19 unit buildings.
- Across ownership, owners are more likely to participate (32% probability) than renters (27% probability).
- Across building age, buildings constructed after 1950 are more likely to participate (31% probability) than buildings built before 1950 (28% probability).

There are also several interesting interactions:

- There is little variation in participation likelihood for the different types of single-family or duplex buildings.
- Within triplexes and fourplexes, building age is more important than ownership status. There is little difference in participation probability for renters versus owners.
- Both ownership and building age make substantial differences for buildings that have 5 or more units. This pattern could represent a compounding of the different barriers to participation faced by renters, and people living in older buildings, along with potential income effects (renters living in older buildings likely have lower incomes than owners living in newer buildings).
- The highest participation probaiblities are among the largest buildings that are owned and built after 1950. These likely represent condos in large buildings, probably concentrated in urban areas and possibly occupied by individuals with greater incomes.

4.6.2.2 Electric savings rate

Table 4-55 shows how often the Monte Carlo ANOVA runs produced different patterns of statistically significant results for electric savings rate. Two results are consistent predictors of the savings rates of participating buildings: building size, and the interaction between building size and ownership. Building size was a significant predictor in 95% of the runs. The interaction of building size and ownership was a significant predictor of savings 18% of the time. The alpha level for the significance tests was 0.05. Therefore, any effect that appears less than 5% of the time is likely due to false positives.

[13] 1 unit, 2 units, 3-4 units, 5-9 units, 10-19 units, and 20+ units

Table 4-55. Frequency of significant prediction of location savings rates – electric

Effect		Percentage of Monte Carlo Runs
Main effects	Building size	95%
	Ownership	7%
	Building age	1%
Inter-actions	Building size * Ownership	18%
	Ownership * Building Age	0%
	Building size * Building Age	7%

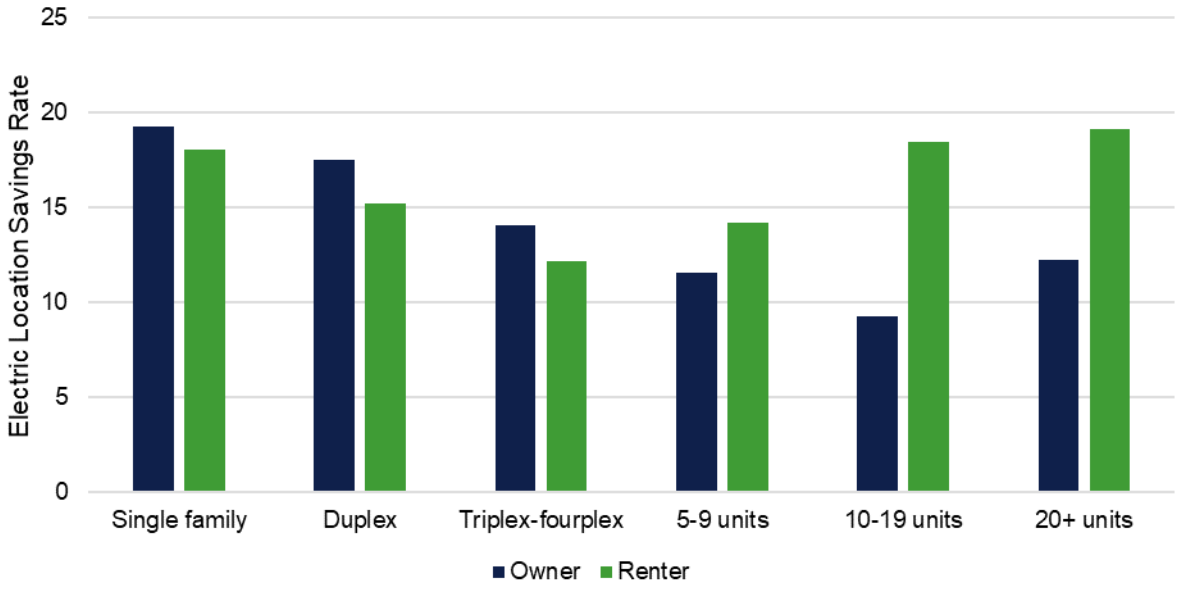
Table 4-56 shows the mean electric location savings rates for each of the 24 categories. These data allow one to interpret the consistently significant effects from the Monte Carlo analysis.

- *Building size main effect:* Participating single-family buildings have the greatest depth of savings, and 5-9-unit buildings have the least. Triplex-fourplex buildings also have a low depth of savings.
- *Building size * Ownership interaction:* Figure 4-27 shows the owner and renter rows of Table 4-56 to explore the nature of the interaction. The figure shows that the difference in savings rates between owners and renters increases as the size of the building increases. There is little difference for single-family buildings, but a large difference for 10-19 and 20+ unit buildings. Interestingly, for the two largest building categories, renters are achieving deeper savings than owners.

Table 4-56. Mean location savings rate by category - electric

	Single family	Duplex	Triplex-fourplex	5-9 units	10-19 units	20+ units	Total
Owners	19.3	17.5	14.0	11.5	9.3	12.2	14.0
Built post 1950	19.9	17.6	13.6	10.3	10.1	11.9	13.9
Built pre 1950	18.7	17.4	14.4	12.8	8.4	12.5	14.0
Renters	18.0	15.2	12.2	14.2	18.4	19.1	16.2
Built post 1950	19.0	15.4	12.1	13.5	15.7	17.9	15.6
Built pre 1950	17.0	14.9	12.3	15.0	21.1	20.4	16.8
Total	18.7	16.4	13.1	12.9	13.8	15.7	15.1

Figure 4-27. Building size * Ownership interaction – electric location savings rate



4.6.2.3 Gas participation rates

The participation analysis estimated the probability of an individual location participating in a gas efficiency program at least once between 2013 and 2019, for each of the 24 location categories (pre/post 195 construction * renter/owner * six building size categories).[14] The probability did not exceed 50% for any category (Table 4-57).

Table 4-57. Location-level probability of participation – gas

	Single family	Duplex	Triplex-fourplex	5-9 units	10-19 units	20+ units	Total
Owners							
Built post 1950	25%	19%	24%	19%	23%	30%	**24%**
Built pre 1950	23%	16%	15%	14%	7%	31%	**21%**
Renters							
Built post 1950	23%	18%	16%	7%	9%	10%	**19%**
Built pre 1950	21%	16%	12%	5%	6%	27%	**16%**
Total	**24%**	**17%**	**15%**	**8%**	**11%**	**14%**	

The main effects are very consistent with the block group-level results:

- Across the building size category, single-family homes had the greatest probability of participating (24%). Five to 9-unit buildings had the least likelihood to participate (8%).
- Across ownership, owners were more likely to participate (23%) than renters (18%).
- Across building age, buildings constructed after 1950 were more likely to participate (22% probability) than buildings built before 1950 (19% probability).

There are also several interesting interactions:

[14] 1 unit, 2 units, 3-4 units, 5-9 units, 10-19 units, and 20+ units

- Within single-family buildings, duplexes, and triplexes/fourplexes, both ownership and building age have small effects. Owners and newer buildings are more likely to participate.
- In 5-9-unit buildings, both ownership and building age make substantial differences. Owners and newer buildings are more likely to participate.
- For 10-19-unit buildings, owners of newer buildings (likely condos) are more likely to participate, and the other four combinations are less likely to participate.
- Within 20+ unit buildings, renters of newer buildings are least likely to participate, and the other three combinations have the greatest participation probabilities across all 24 categories.

4.6.2.4 Gas savings rate

Table 4-58 shows how often the Monte Carlo ANOVA runs produced different patterns of statistically significant results for gas savings rate. Three results are consistent predictors of the gas savings rates of participating buildings: ownership, building size, and the interaction between building size and ownership. Ownership was a significant predictor 65% of the time, building size was a significant predictor in 34% of the runs, and the interaction of building size and ownership was a significant predictor of gas savings 13% of the time. The alpha level for the significance tests was 0.05. Therefore, any effect that appears less than 5% of the time is likely due to false positives.

Table 4-58. Frequency of significant prediction of location savings rate - gas

Effect		Percentage of Monte Carlo Runs
Main effects	Building size	34%
	Ownership	65%
	Building age	3%
Inter-actions	Building size * Ownership	13%
	Ownership * Building Age	2%
	Building size * Building Age	4%

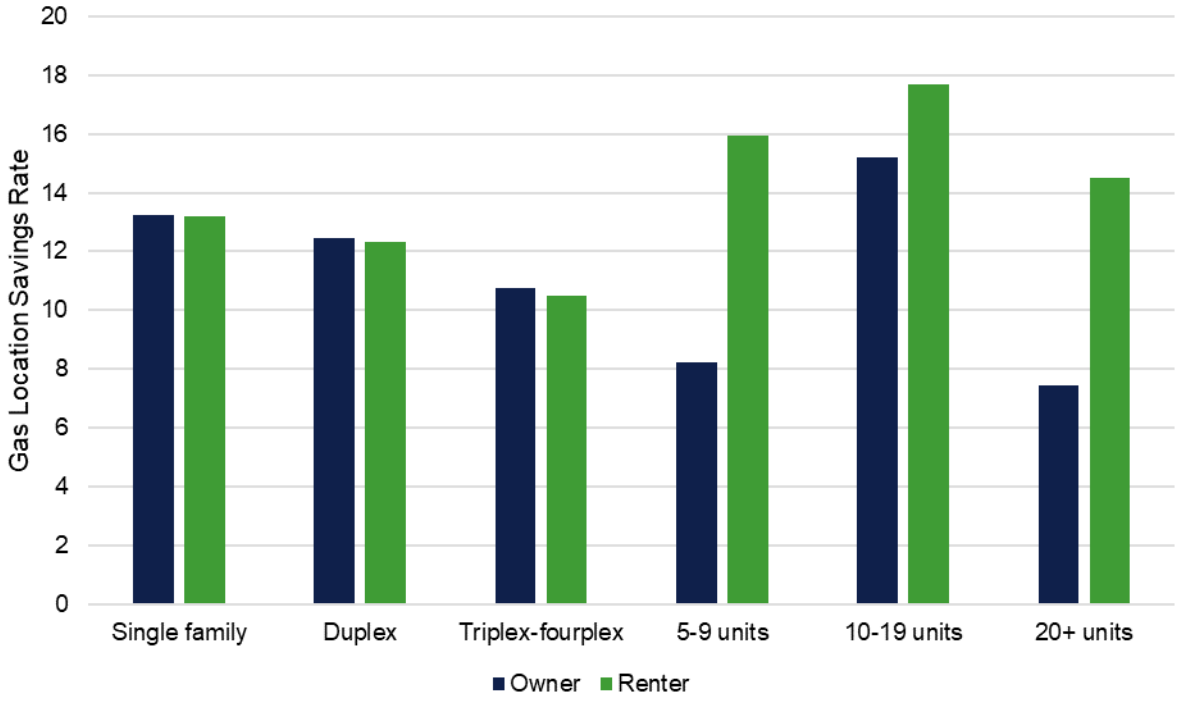
Table 4-59 shows the mean gas location savings rates for each of the 24 categories. This data allows one to interpret the consistently significant effects from the Monte Carlo analysis.

- *Ownership main effect:* Participating renters have a greater depth of savings (savings rate = 14.0) than owners (savings rate = 11.2).
- *Building size main effect:* 10-19-unit buildings have the greatest depth of gas savings (savings rate=16.4), and triplexes/fourplexes have the least (savings rate=10.6).
- *Building size * Ownership interaction:* The ownership main effect appears to be driven by the interaction between building size and ownership (Figure 4-28) There is little difference in gas savings rates for participating buildings that are single family, duplex, or triplex/fourplex. However, for 5+ unit buildings, participating renters have a greater depth of gas savings than owners.

Table 4-59. Mean location savings rate by category - gas

	Single family	Duplex	Triplex-fourplex	5-9 units	10-19 units	20+ units	Total
Owners	13.2	12.5	10.8	8.2	15.2	7.4	11.2
Built post 1950	13.6	12.5	10.9	9.1	7.8	8.7	10.4
Built pre 1950	12.9	12.4	10.6	7.3	22.6	6.1	12.0
Renters	13.2	12.3	10.5	15.9	17.7	14.5	14.0
Built post 1950	13.9	12.7	10.0	15.3	13.4	19.3	14.1
Built pre 1950	12.4	12.0	11.0	16.6	22.0	9.6	13.9
Total	13.2	12.4	10.6	12.1	16.4	11.0	12.6

Figure 4-28. Building size*Ownership interaction – gas location savings rate



5 APPENDICES

5.1 Full correlation matrices

This appendix presents the full correlation matrices for electric and gas. It also includes figures that combine the correlations of all three participation metrics for each fuel type.

Table 5-1. Full correlation matrix - electric

	Savings/Consumption	Consumption-weighted participation	Location-level participation	Low income	Moderate income	Average or higher income	Renter-occupied housing	Owner-occupied housing	Multifamily housing	Single houses	Duplex, triplex, fourplex housing	Limited English-speaking households	Construction year pre-1950	Construction year post 50	Urban PA households ratio
Location-level participation	0.54	0.74	1.00	-0.34	-0.18	0.37	-0.42	0.42	-0.30	0.37	-0.23	-0.23	-0.22	0.22	0.04
Consumption-weighted participation	0.49	1.00	0.74	-0.28	-0.17	0.31	-0.30	0.30	-0.06	0.22	-0.25	-0.18	-0.23	0.23	0.05
Savings/Consumption	1.00	0.49	0.54	-0.24	-0.18	0.28	-0.38	0.38	-0.18	0.38	-0.36	-0.30	-0.33	0.33	-0.04
Urban PA households ratio	-0.04	0.05	0.04	0.12	0.01	-0.10	0.29	-0.29	0.22	-0.32	0.26	0.30	0.25	-0.25	1.00
Construction year post 50	0.33	0.23	0.22	-0.24	-0.12	0.26	-0.49	0.49	-0.09	0.56	-0.70	-0.33	-1.00	1.00	-0.25
Construction year pre-1950	-0.33	-0.23	-0.22	0.24	0.12	-0.26	0.49	-0.49	0.09	-0.56	0.70	0.33	1.00	-1.00	0.25
Limited English-speaking households	-0.30	-0.18	-0.23	0.37	0.14	-0.38	0.58	-0.58	0.30	-0.56	0.51	1.00	0.33	-0.33	0.30
Duplex, triplex, fourplex housing	-0.36	-0.25	-0.23	0.35	0.19	-0.38	0.65	-0.65	0.03	-0.75	1.00	0.51	0.70	-0.70	0.26
Single houses	0.38	0.22	0.37	-0.52	-0.18	0.52	-0.90	0.90	-0.68	1.00	-0.75	-0.56	-0.56	0.56	-0.32
Multifamily housing	-0.18	-0.06	-0.30	0.38	0.05	-0.35	0.65	-0.65	1.00	-0.68	0.03	0.30	0.09	-0.09	0.22
Owner-occupied housing	0.38	0.30	0.42	-0.64	-0.22	0.64	-1.00	1.00	-0.65	0.90	-0.65	-0.58	-0.49	0.49	-0.29
Renter-occupied housing	-0.38	-0.30	-0.42	0.64	0.22	-0.64	1.00	-1.00	0.65	-0.90	0.65	0.58	0.49	-0.49	0.29
Average or higher income	0.28	0.31	0.37	-0.91	-0.53	1.00	-0.64	0.64	-0.35	0.52	-0.38	-0.38	-0.26	0.26	-0.10
Moderate income	-0.18	-0.17	-0.18	0.14	1.00	-0.53	0.22	-0.22	0.05	-0.18	0.19	0.14	0.12	-0.12	0.01
Low income	-0.24	-0.28	-0.34	1.00	0.14	-0.91	0.64	-0.64	0.38	-0.52	0.35	0.37	0.24	-0.24	0.12

Table 5-2. Full correlation matrix - gas

	Savings/Consumption	Consumption-weighted participation	Location-level participation	Low income	Moderate income	Average or higher income	Renter-occupied housing	Owner-occupied housing	Multifamily housing	Single houses	Duplex, triplex, fourplex housing	Limited English-speaking households	Construction year pre-1950	Construction year post 50	Urban PA households ratio
Location-level participation	0.59	0.89	1.00	-0.43	-0.26	0.47	-0.42	0.42	-0.25	0.36	-0.27	-0.26	-0.17	0.17	-0.04
Consumption-weighted participation	0.58	1.00	0.89	-0.43	-0.24	0.46	-0.40	0.40	-0.20	0.33	-0.27	-0.28	-0.15	0.15	-0.04
Savings/Consumption	1.00	0.58	0.59	-0.27	-0.16	0.30	-0.33	0.33	-0.17	0.33	-0.29	-0.24	-0.25	0.25	-0.02
Urban PA households ratio	-0.02	-0.04	-0.04	0.12	0.06	-0.13	0.22	-0.22	0.15	-0.24	0.21	0.21	0.21	-0.21	1.00
Construction year post 50	0.25	0.15	0.17	-0.24	-0.13	0.25	-0.48	0.48	-0.07	0.55	-0.69	-0.32	-1.00	1.00	-0.21
Construction year pre-1950	-0.25	-0.15	-0.17	0.24	0.13	-0.25	0.48	-0.48	0.07	-0.55	0.69	0.32	1.00	-1.00	0.21
Limited English-speaking households	-0.24	-0.28	-0.26	0.39	0.17	-0.40	0.57	-0.57	0.27	-0.55	0.50	1.00	0.32	-0.32	0.21
Duplex, triplex, fourplex housing	-0.29	-0.27	-0.27	0.38	0.21	-0.41	0.67	-0.67	0.04	-0.77	1.00	0.50	0.69	-0.69	0.21
Single houses	0.33	0.33	0.36	-0.54	-0.22	0.54	-0.90	0.90	-0.66	1.00	-0.77	-0.55	-0.55	0.55	-0.24
Multifamily housing	-0.17	-0.20	-0.25	0.38	0.09	-0.36	0.64	-0.64	1.00	-0.66	0.04	0.27	0.07	-0.07	0.15
Owner-occupied housing	0.33	0.40	0.42	-0.65	-0.25	0.65	-1.00	1.00	-0.64	0.90	-0.67	-0.57	-0.48	0.48	-0.22
Renter-occupied housing	-0.33	-0.40	-0.42	0.65	0.25	-0.65	1.00	-1.00	0.64	-0.90	0.67	0.57	0.48	-0.48	0.22
Average or higher income	0.30	0.46	0.47	-0.92	-0.54	1.00	-0.65	0.65	-0.36	0.54	-0.41	-0.40	-0.25	0.25	-0.13
Moderate income	-0.16	-0.24	-0.26	0.16	1.00	-0.54	0.25	-0.25	0.09	-0.22	0.21	0.17	0.13	-0.13	0.06
Low income	-0.27	-0.43	-0.43	1.00	0.16	-0.92	0.65	-0.65	0.38	-0.54	0.38	0.39	0.24	-0.24	0.12

Figure 5-1. Correlation of participation metrics with demographic variables – electric

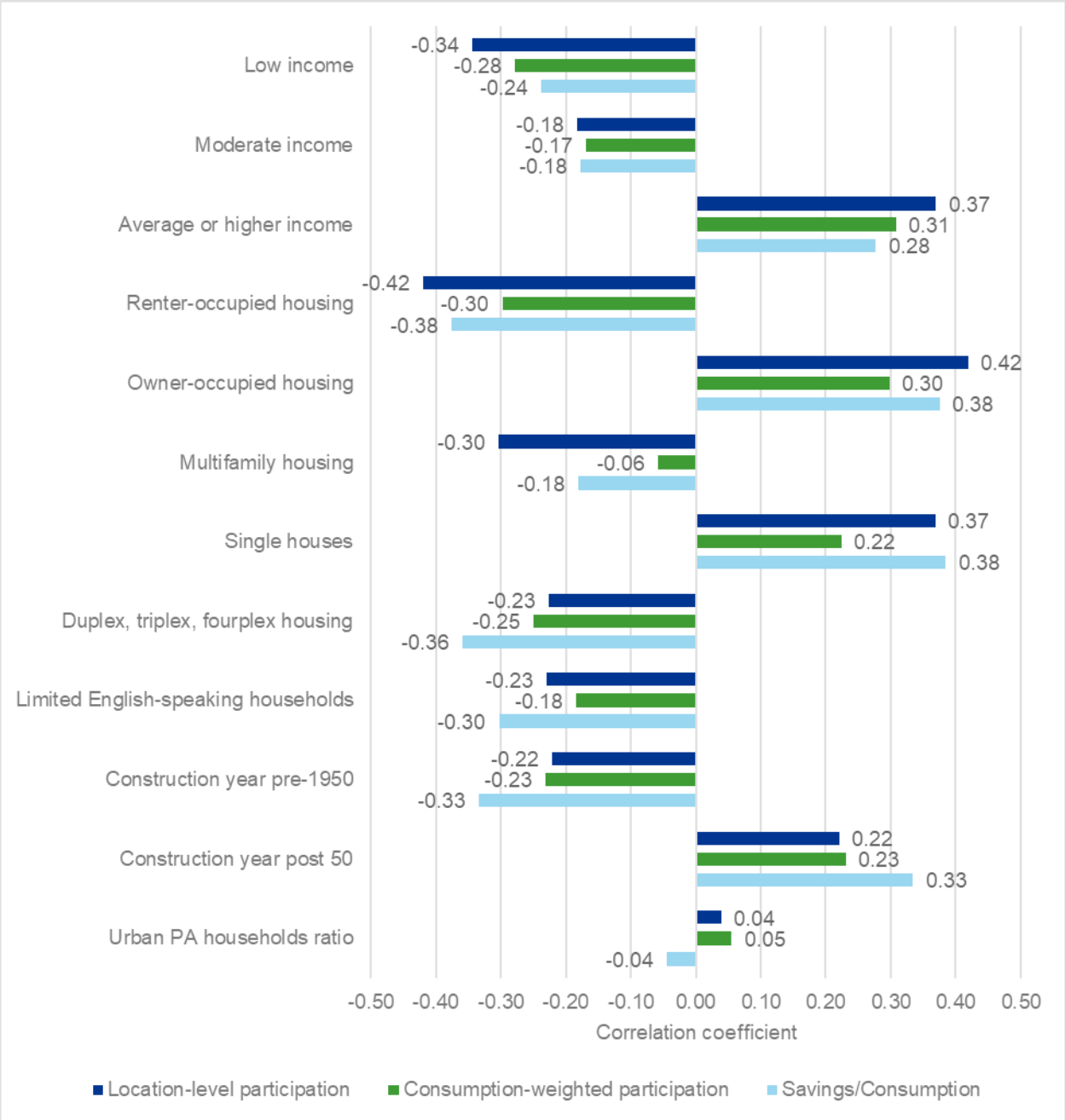
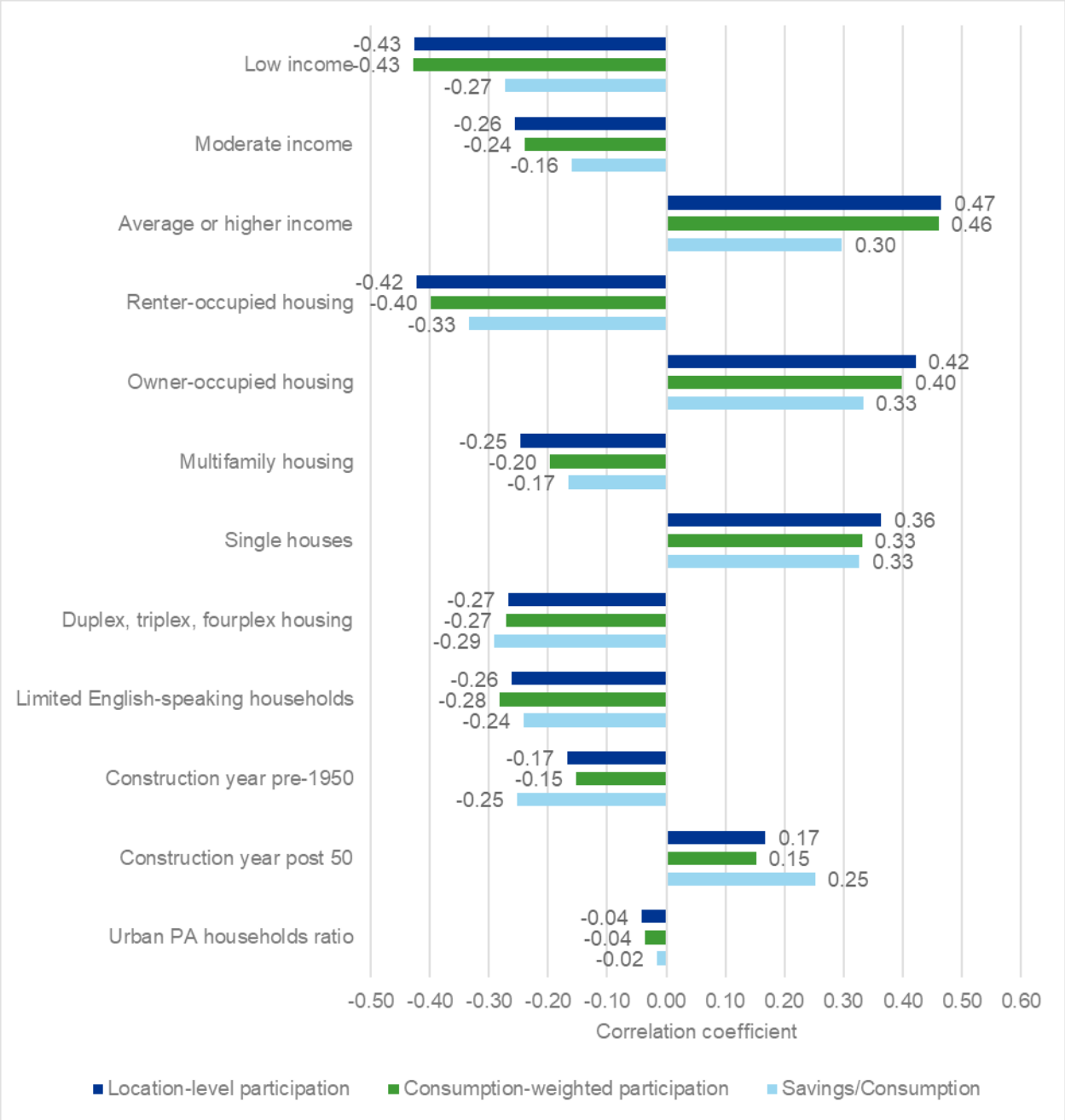


Figure 5-2. Correlation of participation metrics with demographic variables – gas



5.2 Input value definitions

Below we define the various input values used in the RCPS. Wherever possible, the appendix uses terms as defined by the MA PAs; where appropriate, the definitions are also consistent with those used in the C&I Customer Profile Study. Sources of these definitions include the Mass Save Data website, the MA 2019-2021 Three-Year Plan, the MA technical reference manual (TRM), and the PAs' benefit-cost screening models.

Account: The unique combination of customer and premise (defined below).

Consumption: An account's energy use over a given period, measured in kilowatt-hours (electric) or therms (gas). Most often in the RCPS, "consumption" refers to annual consumption for either an account, a location, or an aggregation of these to the block group level.

Gross kWh/therm savings: Throughout this report, "savings" refers to gross annual energy savings (in kWh or therms), as provided by the PAs in their tracking data. These savings are calculated with measure algorithms and do not include impact factors (i.e., in-service rate, savings persistence factor, realization rate, and summer and winter peak demand coincidence factors).

Incentives: Per the MA TRM, incentives are "funds paid by the reporting Program Administrator to or on behalf of customers or trade allies as rebates or in other forms."

Location: A unique standardized physical address in the PAs' data. A 3-unit apartment complex at "1 Main St, Anytown, MA 12345" would be classified as a single location.

Participant: Per the MA TRM, a participant is "a customer who installs a measure through regular program channels and receives any benefit (i.e., incentive) that is available through the program because of their participation." The RCPS defines participants based on either unique accounts or unique locations depending on the analysis grain.

Premise: The physical location of an account. A 3-unit apartment complex at "1 Main St, Anytown, MA 12345" would be classified as 3 premises. A single-family home would be classified as one premise.

Quantity: Typically, the number of measures an energy efficiency participant installs. Sometimes quantity is not just the number of efficient fixtures/bulbs installed but can be square feet of insulation or hours completed by a contractor to perform an assessment or installation. The RCPS does not typically report on quantity values contained in the MA Evaluation Database due to the variation of units and meaning behind those units for select energy efficiency activities.

Tracking account: The unique combination of a customer and premise that participated in an energy efficiency program.

Tracking location: A unique standardized physical address in the PAs' energy efficiency program tracking data.

5.3 Metrics definitions

DNV generated the following metrics as part of the previous profile reports for readability, we use the term *analysis grain* in these equations instead of writing out "*accounts and/or premises*":

Participant savings achieved: This statistic looks at the energy savings from participant accounts/premises relative to the consumption for those participants only. It can answer questions such as, "How much energy did participating residential accounts/premises save relative to their consumption?"

$$\frac{(Sum\ of\ savings\ for\ participating\ [analysis\ grain]\ within\ analysis\ population)}{(Sum\ of\ consumption\ of\ participating\ [analysis\ grain]\ only\ within\ analysis\ population)}$$

Population consumption-weighted participation: This statistic looks at the consumption of efficiency participants within a specific analysis grain relative to the consumption of the total analysis grain. It can answer questions such as, "What percentage of total residential consumption for each PA participated in an efficiency program in 2016?"

$$\frac{(Sum\ of\ consumption\ for\ participating\ [analysis\ grain]\ within\ analysis\ population)}{(Sum\ of\ consumption\ of\ all\ [analysis\ grain]\ within\ analysis\ population)}$$

Population participation: This statistic identifies the ratio of accounts/premises within the analysis population that participated in energy efficiency programs. It can answer questions such as, "What percentage of identified residential accounts/premises participated in an efficiency program for each PA?"

$$\frac{(Number\ of\ participating\ [analysis\ grain]\ within\ analysis\ population)}{(Number\ of\ [analysis\ grain]\ within\ analysis\ population)}$$

Population savings achieved: This statistic looks at the energy savings of efficiency participants within a specific analysis grain relative to the consumption of the total analysis grain. It can answer questions such as, "How much energy did the residential sector save relative to its total consumption?"

$$\frac{(Sum\ of\ savings\ for\ participating\ [analysis\ grain]\ within\ analysis\ population)}{(Sum\ of\ consumption\ of\ all\ [data\ grain]\ within\ analysis\ population)}$$

Proportional contribution ratio: This statistic looks at the percentage of total savings within an analysis bin divided by the percentage of total consumption within the same analysis bin. This lens provides a more normalized view of a bin's savings relative to consumption by incorporating population-level weights. It can answer questions such as, "Was residential accounts' savings contribution proportional to their consumption contribution by analysis bin?"

$$\frac{\left(\frac{Savings\ for\ analysis\ bin}{Total\ savings\ for\ all\ analysis\ bins}\right)}{\left(\frac{Consumption\ for\ analysis\ bin}{Total\ consumption\ for\ all\ analysis\ bins}\right)}$$

Total populations savings: This statistic represents the cumulative savings over all years that a measure has been installed. (Note that this calculation may lead to an overstatement of savings if any installed measures had a life of fewer than 7 years.) In section 4.1.2, total population savings were calculated using the following formula:

*(2013 Non-behavioral Savings * 7) + (2014 Non-behavioral Savings * 6) + (2015 Non-behavioral Savings * 6) + (2016 Non-behavioral Savings * 4) + (2017 Non-behavioral Savings *3) + (2018 Non-behavioral Savings *2) + (2019 Non-behavioral Savings *1) + (2013, 2014, 2015, 2016, 2017, 2018 and 2019 Behavioral Savings)*

5.4 RCPS analysis grains definitions

This appendix defines the various analysis grains used in the RCPS.

ACS block group: Per the Federal Register, "Block groups are statistical geographic subdivisions of a census tract and are the smallest geographic areas for which the Census Bureau provides sample data, primarily from the ACS 5-year period estimates...a block group should contain either at least 600 people or at least 240 housing units at minimum, and 3,000

people or 1,200 housing units at maximum." The RCPS uses block groups to display geographic data, and to provide underlying information for tabular and graphical outputs of demographic groupings.

Core initiative: A channel within a program (defined below) that includes a set of efficiency offerings tailored to various program sub-groups. For the 2019 RCPS, the Residential core initiatives are:

2016-2018 Core Initiative	2019-2021 Core Initiative
New Construction	Residential New Homes & Renovations
Multi-Family Retrofit	Residential Coordinated Delivery
Home Energy Services - Measures	Residential Coordinated Delivery
Behavior/Feedback Program	Residential Behavior
Heating & Cooling Equipment	Residential Retail
Consumer Products	Residential Retail
Lighting	Residential Retail
Single Family Retrofit	Income Eligible Coordinated Delivery
Multi-Family Retrofit	Income Eligible Coordinated Delivery

Customer type: The customer category of "residential" or "low-income"

End use: Per the MA TRM, end use refers to "the category of end use or service provided by a measure or technology (e.g., lighting, cooling, etc.)."

Program: The broad channel through which accounts participate in energy efficiency. Per the MA benefit-cost screening models, the Residential programs are Low-Income Whole House, Residential Products, and Residential Whole House. Each program contains at least one core initiative (defined above).

Use code: Three-digit code used to classify how a particular property is used (e.g., single-family residence, mobile home residence)