

Prepared by (in alphabetical order)

Kristen Bush, Jessica Dunn, Sadia Gul,   
Mark T. Lozano, Leslie Nelson, Tobiah Steckel

An Analysis of Residential underserved customers in California Investor-Owned Utility Programs

Prepared for: the CAEECC sub-Working Group on Underserved Customers

Under the supervision of Prof. Alissa Kendall (amkendall@ucdavis.edu), and with the Assistance of Dr. Hal Nelson

## Contributions of the authors:

K. Bush organized and conducted the literature review of equity and energy justice in relation to residential energy efficiency programs and contributed to the writing and editing of the document.

J. Dunn researched statistical methods used to review energy efficiency programs, carried out the multivariate regression, summarized results, and provided discussion of key findings.

S. Gul located the relevant literature and screened it, helped developed spatial figures using QGIS, and contributed to the editing of the document.

M. T. Lozano helped develop spatial figures using QGIS, investigated the equity implications of the considered programs by comparing the distribution of benefits to the affected communities’ CalEnviroScreen burden scores, and contributed writing and editing to the document.

L. Nelson reviewed program implementation plans to add additional program information data, provided expertise on programs terminology and classification, reviewed statistical results, and provided recommendations for future data gathering.

T. Steckel conceptualized analyses, generated datasets and carried out formal analysis of t-tests between bottom and top quintile.

## Acknowledgements:

Thanks to Dr. Hal Nelson for providing guidance on the statistical methods and approached used in this analysis; and Lara Ettenson of NRDC for provision of data and bringing the project to this team. This research was completed as part of a graduate class at University of California Davis entitled: EGG298A – Data Science for Equity in Energy Systems (Instructor: Prof. Alissa Kendall)

Table of Contents

[Acknowledgements and contributions of the authors: 1](#_Toc59128719)

[1. Introduction 4](#_Toc59128720)

[1.1 Motivation 4](#_Toc59128721)

[1.2 Research Questions 4](#_Toc59128722)

[2. Literature Review 4](#_Toc59128723)

[2.1 Criteria for Literature Review 5](#_Toc59128724)

[2.2 Defining Energy Justice and Equity 5](#_Toc59128725)

[2.3 Examining the Implementation of Residential Energy Efficiency Programs 7](#_Toc59128726)

[2.4 Literature Review Outcomes 9](#_Toc59128727)

[2.5 Suggestions for Improving Residential Energy Efficiency Programs 9](#_Toc59128728)

[3. Methods 10](#_Toc59128729)

[3.1 Review of Methods Used in Previous Assessment of Inequity in Energy Efficiency Programs 10](#_Toc59128730)

[3.2 Methods Selected for Analysis 10](#_Toc59128731)

[3.3 Other Methods Explored 11](#_Toc59128732)

[3.4 Data Sources 11](#_Toc59128733)

[4. Results 13](#_Toc59128734)

[4.1 Zip Code Level Analyses 13](#_Toc59128735)

[4.2 Program-Filtered Analyses 23](#_Toc59128736)

[5. Discussion and Conclusions 23](#_Toc59128737)

[5.1 Spatial Distribution 23](#_Toc59128738)

[5.2 What the analyses suggest about underserved populations 24](#_Toc59128739)

[5.3 The effect program type on highest and lowest quintiles 24](#_Toc59128740)

[5.4 Analytical limitations resulting from zip code level aggregation of claims data 25](#_Toc59128741)

[5.5 Limitations and suggestions for CEDARS program data 25](#_Toc59128742)

[Works Cited 27](#_Toc59128743)

[Appendix 29](#_Toc59128744)

[Section A1. Explanatory Variables 29](#_Toc59128745)

[Section A2. Multivariate regression results 29](#_Toc59128746)

[Section A3. Additional spatial mapping figures 31](#_Toc59128747)

Table of Tables

[Table 1. Summary statistics for explanatory variables across all zip codes. 14](#_Toc59117394)

[Table 2 reports a summary of the population-weighted average CalEnviroScreen percentile burden score for the upper and lower quintiles for each tracked variable. Note that, on average, communities who see the least benefits from energy efficiency programs tend to have higher burden scores than those who see the most benefits 19](#_Toc59117395)

[Table 2. Average population-weighted CES percentile burden score for the lower and upper quintiles of response variables. 19](#_Toc59117396)

[Table 3. T-test results for the bottom quintile, top quintile, and p-value of the mean difference between quintile populations (lq = lower quintile, uq = upper quintile). 20](#_Toc59117397)

[Table 4. Multivariate regression results that have statistical significance for at least one of the response or explanatory results.. 22](#_Toc59117398)

[Table A1. Complete multivariate regression results. The coefficient is listed first, followed by the standard errors in parentheses. \*, \*\*, \*\*\* indicates significance at the 90%, 95%, and 99% level, respectively. 29](#_Toc59117399)

Table of Figures

[Figure 1. Frequency of zip codes in bottom quintile for response variables (participation rate, lifecycle net kilowatt-hour (kWh) savings per household, lifecycle net therm savings per household, gross incentive per household, and end user rebate per household). 15](#_Toc59094511)

[Figure 2. Zip codes in the lowest and highest quintile for gross incentive per household for all claims (2017-2019). 16](#_Toc59094512)

[Figure 3. Zip codes in the lowest and highest quintile for total net kWh savings per household for all claims (2017-2019). 16](#_Toc59094513)

[Figure 4. Zip codes in the lowest and highest quintile for total net therm savings per household for all claims (2017-2019). 17](#_Toc59094514)

[Figure 5. Zip codes in the lowest and highest quintile for the total unit rebate per household for all claims (2017-2019). 17](#_Toc59094515)

[Figure 6. Zip codes in the lowest and highest quintile for the number of units installed per household for all claims (2017-2019). 18](#_Toc59094516)

[Figure A1. CalEnviroScreen percentile burden scores. Those with scores in the upper quartile (dark red) are considered disadvantaged communities. 31](#_Toc59094517)

[Figure A2. The mean total net KWh savings of energy efficiency programs across different climate zones. 32](#_Toc59094518)

[Figure A3. The mean total net therm savings of energy efficiency programs across different climate zones. 33](#_Toc59094519)

[Figure A4. The distribution of the energy efficiency programs compared to the location of urban areas in California. 34](#_Toc59094520)

## 1. Introduction

### 1.1 Motivation

Energy efficiency has been touted as a relatively simple way of contributing to state climate and energy goals, especially in terms of costs to program administrators when compared to other means of climate change mitigation. In particular, residential energy efficiency programs have become increasingly prevalent as program administrators (i.e., investor-owned utilities (IOUs) and regional energy networks (RENs)) continue to pursue their implementation. This growth is fitting as residential energy consumption accounts for approximately 40% of U.S. energy consumption (Sokolova et al., 2019), and 19% of California’s energy consumption (U.S. Energy Information Administration, 2020). Although the growth of these programs can be viewed as a positive reflection of the commitment made to meeting state climate and energy goals, the uptick in residential energy efficiency programs also calls for a more thorough evaluation of whether they are equitably deployed.

Under the direction of the California Public Utilities Commission (CPUC), working groups such as the California Energy Efficiency Coordinating Committee (CAEECC) provide oversight of energy efficiency programs to ensure that the needs of underserved customers and households (e.g., multi-family, renter-occupied, low- and moderate-income, etc.) are met while also being mindful of market dynamics. However, more work is required to understand who is underserved (i.e., those who are not benefitting from energy efficiency programs) and which barriers cause different segments of customers to remain underserved despite programs having good intentions.

### 1.2 Research Questions

CAEECC seeks to better understand which residential customers are underserved to properly increase the availability and accessibility of energy efficiency programs for those customers, with an emphasis on customers who are experiencing a high energy burden but face challenges in program eligibility. Further, there are concerns that customers who do not qualify for low income programs, like the Energy Savings Assistance Program, but are in lower income brackets may not be receiving sufficient energy efficiency program benefits. These concerns have been translated into the following research questions, which have guided the analysis detailed in this report:

What are the conceptual models that have been proposed to understand inequities and injustices in the energy system and energy efficiency programs in particular?

Are there differences in access to the residential energy efficiency program benefits across geographic areas and socio-demographic groups in California?

Do certain program characteristics improve service?

Do certain program characteristics provide more equitable access?

The research undertook both a comprehensive literature review, as well as statistical and spatial data analyses.

## 2. Literature Review

The goal of the literature review was to identify the conceptual models that have been proposed to understand (1) inequities and injustices in both energy system and energy efficiency programs, and (2) methods previously used to assess the inequitable distribution of the outcomes and services of energy efficiency programs based on concepts of energy justice and equity. This review provides additional context to better understand the reasons behind these disparities and the methods used to evaluate them. By identifying previous methods and perspectives for understanding spatial and socio-demographic inequities in program deployment, this literature review informs the spatial and statistical analyses conducted as part of this research project.

### 2.1 Criteria for Literature Review

Keywords used in collecting the literature included: energy efficiency inequality, median income, Tony Reames, environmental justice, and Liebert.

One method of collecting the literature used in this literature review included utilizing the keywords listed above to search for relevant articles and papers contained within Google Scholar. This search was filtered to exclude articles and papers that were published before 2016.

Another method of collecting literature considered factors affecting energy consumption before moving to energy efficiency programs and equity. After a relevant paper was found, a search for papers citing it was conducted. This was done for subsequent papers as well.

### 2.2 Defining Energy Justice and Equity

#### 2.2.1 Conceptualizations of Energy Justice

Conceptualizations of energy justice acknowledge that environmental and social inequities are commonly derived from and reflected in discrepancies in energy consumption. Differences in energy consumption are primarily driven by human behavior, which is in turn driven by spatial and socio-economic factors, including infrastructure, economics, and culture. Hence, it is important to understand the correlation between behavior, energy conservation, and consumption, and the effect they have on the implementation and design of energy policies (Elnakat et al., 2016). Ignoring these differences impedes the progress necessary to achieve the basic human right to energy, which includes: “the right to a healthy, sustainable energy production; the right to best available energy infrastructure; the right to affordable energy; and the right to uninterrupted energy service” (Reames, 2016b). These differences, or ignorance of these differences, perpetuates issues related to the equitable distribution of energy services including inequalities in technology energy efficiency, energy burden, and energy prices. Inclusive participation in energy conservation and related activities requires affordable and socially accepted energy-efficient technologies, programs, and policies that can be accessed by the most marginalized communities. Not only does this manifest itself differently between households, but also within households as well (Reames et al., 2018).

Energy justice frameworks often emphasize the importance of social practices and relations that inform energy use and decision-making in the context of “community.” This includes an awareness of the broader social context that leaves people culturally and economically place-bound and how this results in identifiable clusters of social deprivation. By accounting for the differences in the lived experiences of and the challenges that need to be overcome by different communities, improvements to both energy justice and equity can be made without having to rely on broad and homogenous program implementation (Reames, 2016a). This would also assist in ushering in an equitable and just energy transition that pushes for affordable energy for all, evenly distributed benefits, and sustainable technologies. The concept of just energy finance has also been condensed into the following principles: affordability, good governance, due process, intra-generational equity, spatial equity, and finance resilience (Forrester, 2020).

Additionally, the linkages between energy justice and equity can be considered through the lens of procedural equity, which refers to the transparency and fairness of processes that allocate resources, resolve conflicts, and divide benefits and burdens. Other forms of equity include: distributive equity, (i.e., access) which limits the influence of privileges and prioritizes those with the highest need; intergenerational equity, which considers obligations to future generations; redistributive equity, which involves fairness in the punishment of wrongs; structural equity, which prevents cumulative disadvantage felt by subordinated groups; and transgenerational equity, which is the avoidance of unfair burdens on future generations (Brown et al., 2020).

#### 2.2.2 Barriers to Energy Justice

Energy justice encompasses all the barriers that prove detrimental to the quality of life, well-being, and autonomy and seeks to bring about actionable solutions to counter them holistically. Of the existing barriers, energy poverty, fuel poverty and energy burden are the most detrimental and deeply intertwined with socio-demographic traits such as race, income, and education. An informed view is required to understand the ways in which these barriers negatively affect the accessibility and affordability of energy services and products and allow for vulnerabilities in consumer finances, health, and location to manifest (Ambrose et al., 2019).

Energy poverty connects challenges related to energy inequality (i.e., energy injustice) and energy justice and highlights the particular barriers for low-income households (Xu et al., 2019). Energy poverty often arises from a combination of energy-inefficient housing, inefficient heating appliances and systems, low income, and high fuel costs. Because it derives from these multitudes of causes, a measurable definition is not available. The effects of energy poverty may include accumulated debt, poor indoor air quality, and temperatures that negatively affect household health (Ambrose et al., 2019, Reames, 2016a). Although energy poverty and fuel poverty are often used interchangeably, proposed definitions have been created to distinguish the two terms. Energy poverty is seen to occur within households that rely only on electricity and gas whereas fuel poverty is a result of a broader collection of energy sources. Apart from economics, fuel poverty also refers to the lived experiences of the fuel poor. Although fuel poverty is a symptom of distributional injustice (i.e., the rejection of the idea that all of society has a right to equal treatment and a fair distribution of outcomes), its prominence is also a result of a broader inability to recognize the energy needs of vulnerable populations and the procedural injustice that keeps them from having a significant role in decision-making, access to information, and access to legal processes needed to challenge unfavorable decision-making processes (Reames, 2016b).

Fuel poverty can also be measured in terms of energy burden, or the proportion of household expenditures used to cover home energy costs. In this case, the share of income used for utility-related expenses is often disproportionate and unaffordable, exceeding 6-11% of a household’s annual gross income according to the U.S. Department of Health Services, and often with unbalanced impacts on minority and low-income households (Xu et al., 2019). This also affects mobile homes and households in rural areas, Indian reservations, island territories as well as households with children, elderly residents, and disabled occupants that have higher energy costs. This highlights the common correlates and causes of energy burden which include behavioral factors, location and geography, socio-economic characteristics, housing characteristics, and energy prices and policies (Brown et al., 2020). The aforementioned metric is based on the premise that a household should allocate no more than 30% of its income on housing expenses, and that utility costs should not be more than 20% of these expenses. However, this has been parsed further to distinguish between “energy stressed” households (4-7%), “energy burdened” households (7-10%), and “energy impoverished” households (10%+) by some scholars (Brown et al., 2020).

Analyses of energy burdens are often incomplete; while they do include total household spending on energy bills (e.g., home energy services, heating, cooling, etc.), they often do not account for spending on transportation energy. Household budget and income are also represented by different markers of poverty and wealth, such as the Supplemental Poverty Measure (SPM), State Median Income (SMI), Federal Poverty Level (FPL), and Area Median Income (AMI). These various definitions can lead to different valuations of energy burden which can consequently influence the energy savings of a program (Brown et al., 2020). As such, residential energy efficiency programs face an arduous task in that they not only have to provide adequate relief from energy burden, energy insecurity, energy access, and energy poverty but that the methods they implement must also include solutions to these challenges based on the various lived experiences of their target customers.

### 2.3 Examining the Implementation of Residential Energy Efficiency Programs

#### 2.3.1 Barriers to Energy Justice in Residential Energy Efficiency

Increasing residential energy efficiency is understood to be an effective and strategic method of eradicating fuel poverty as it can deliver both private benefits, as well as considerable public benefits including avoiding emissions of toxins and pollutants, reduction of power system costs, and the lessening of grid congestion (Reames, 2016b; Zimring et al., 2011). And as such, financing and funding models of residential energy efficiency models have improved and become more innovative (Forrester et al., 2020). However, there are multiple barriers that continue to keep these programs from reaching effectiveness goals.

The first of these barriers is financing. Both traditional and specialized energy efficiency financing products are flawed. While a host of financial products exist that span secured and unsecured loans, short and long term loans, and those tailored to energy efficiency investments (including on-bill financing and repayment) or general loans, many residents lack capital and are not in a position to take on the risk involved with financing energy efficiency improvements. Even those financial products that have been tailored to address this issue, such as savings-backed arrangements (i.e., Managed Energy Service Agreements (MESA), Energy Service Agreements (ESA), and Energy Savings Performance Contracts (ESPC)), and intended to overcome barriers to efficiency investments, are complex to consumers and have not accounted for significant investment volumes (Leventis et al., 2016).

The second of these barriers is the estimation of energy consumption. Many residential energy efficiency programs rely solely on mathematical models to represent energy consumption by a household, assuming residential energy consumption can be determined primarily by household size. This type of analysis reduces or omits the impact of variables such as house/building type, income, geographic location, and personal preferences and/or household practices related to the frequent use of appliances, despite that the commonly used variable of household size typically accounts for just 26.55% of electricity consumption (Sokolova et al., 2019). It also fails to account for the “energy efficiency financing coverage gap” or the difficulty of adequately engaging middle-income households that do not qualify for the benefits offered by low-income programs but cannot make improvements on their own and lack the capital or credit required to do so. This prevents achieving better realized savings and program penetration when targeting underserved markets (Forrester et al., 2020).

A third barrier to the effectiveness of residential energy efficiency programs is the “energy efficiency gap.” While historically definitions of this gap have focused on energy efficiency opportunities that are possible but not cost effective to implement, Reames (2016a) offers a more nuanced definition, where the energy efficiency gap can be conceptualized as either the split between inherent energy efficiency benefits and the actual benefits that have been realized or a prolonged rate of energy efficiency technology development. This gap manifests itself in the form of regulatory barriers (e.g., policies informing eligibility requirements), social barriers (e.g., public priorities and public distrust of government officials, contractors, and technology), and market barriers (e.g., information gaps and split-incentives faced by renters and landlords). These barriers reduce the ability of middle-income households to invest in energy efficiency as they may not be able to address expensive maintenance and structural problems that are necessary prior to the investment. Researchers have also called for program administrators to support the development of the institutional capabilities necessary for programs to meet the specific needs of target groups through flexible policy adjustments (Reames, 2016a). Similarly, energy efficiency portfolios that typically rely on per capita investments, maximization of savings per property, and energy savings per participant lack the structures and systems needed to allow for distributional equity and justice across socioeconomic groups and service territories. This may be exacerbated by program spending requirements/investment levels and income qualifiers (i.e., SPM, SMI, FPL, AMI) that possibly indicate inequities in savings benefits (Reames et al., 2019).

Finally, the proprietary nature of the household and individual data related to energy use intensity also presents a challenge to adequate targeting of residential energy efficiency programs as this lack of accessible data leads to a deficient understanding of the spatial distribution of vulnerability and energy consumption needed to target those that are least advantaged (Reames, 2016b). This may be preceded by a lack of utility interest in programs that can potentially reduce revenue and concerns for customer privacy. Without the implementation of better targeted policies and initiatives, broad market-based interventions (e.g., tax rebates and low-interest loans) will continue to limit the participation of low- and moderate-income households in vulnerable communities and overshadow concerns of social equity and energy justice. This joining of socioeconomic and demographic data with energy consumption data and housing characteristics could offer practical applications for urban and regional planners, electric utilities, and the public (Elnakat et al., 2016).

### 2.4 Literature Review Outcomes

The literature review reveals the gaps that exist between the social characterization of residential energy efficiency programs and the traditional quantitative evaluations of these programs performed by program administrators. The literature review also offers various conceptualizations of energy justice and demonstrates how these concepts have been used to evaluate energy efficiency programs and the barriers to entry that inhibit program participation. Finally, the review suggests that the availability of data may be a barrier to the development, implementation, and evaluation of programs targeting underserved customers or seeking to address inequities.

### 2.5 Suggestions for Improving Residential Energy Efficiency Programs

Although there are many critiques of residential energy efficiency programs, the literature also provides a means for improving them. A major theme related to energy justice refers to an interdisciplinary approach to program implementation that encompasses solutions to both social and technological issues through engineering and building design, behavioral science, urban planning, health, and social disciplines. An emphasis on a wider array of demand response programs that avoid exacerbating existing disadvantages and acknowledge psychological and social motivators of adoption will also prove beneficial for the participation of underserved communities. This should also include effective communication and trust-building on the part of government officials, utilities, program administrators, and other relevant parties (Xu, 2019).

Reductions to participant risks and costs could be facilitated by offering flexible loan terms based on the performance of a project and achieved savings, piloting performance guarantees to assess their impact on demand and household behavior. Leveraging the knowledge and experience of local organizations, community leaders, and peers would also help to frame energy efficiency upgrades in a way that is directly related to and cognizant of household priorities, concerns, and non-energy issues. Cost-effectiveness screenings and tests with input factors such as social discount rates may also assure that non-energy benefits related to safety, equity, public health, and economic development are explicitly considered and/or quantified within policy goals, minimum energy performance standards. and resulting program portfolios (Zimring et al., 2011). The technological scope of residential energy efficiency programs could be expanded by the further implementation of access to distributed energy resources such as smart meters, storage, solar, and electric vehicles guided by the broadening of regulatory and financing options that reflect an awareness of housing conditions, a need for effective communication that is inclusive of language problems, and methods to relieve the landlord/tenant split incentive, and expand access to lower credit ratings and capital to minimize costs and maximize benefits (Brown et al., 2020).

## 3. Methods

### 3.1 Review of Methods Used in Previous Assessment of Inequity in Energy Efficiency Programs

Previous work addressing similar research questions has used linear regression, multivariate regression, and logit regression to assess the equity of energy use and energy programs, with several studies using a combination of these methods (Reames et al., 2018; Dolsak et al., 2020). Reames et al. (2018) performed linear regression to demonstrate the unequal price distribution of LED light bulbs, along with logistic regression to assess the unequal probability of their availability in poor and disadvantaged communities (e.g., in local stores). Both analyses demonstrate differences across poverty strata and store types.

Multivariate regression includes multiple explanatory variables and is often used to identify the impact of socioeconomic characteristics on household energy efficiency (Reames et al., 2016b), energy security (Hernandez et al., 2016), and factors leading to choosing an energy efficiency program (Forrester and Reames, 2020).

### 3.2 Methods Selected for Analysis

This analysis uses a combination of multivariate regression, t-tests, and spatial analysis to identify underserved populations, based on both the zip code-level socioeconomic data and the program characteristics.

First, the multivariate regression includes the interaction between socioeconomic variables and identifies those variables that are most likely to lead to being served, or not served, by an energy efficiency program. The explanatory variables included factors related to race, income, and housing type (see Appendix Section A1 for the full list).

The regression was run using several different response variables to capture the multiple ways of measuring a household as being “served” or “not served” by energy efficiency programs. For example, “serving” a resident can be a function of the relative energy savings, dollar savings, and/or the money spent by the program administrator on a program to increase these savings for a given resident. In an attempt to capture this nuance, the following response variables were analyzed:

**Participation rate:** The number of program units delivered per household

**Lifecycle net kilowatt-hour (kWh) savings per household:** The total lifecycle net kWh savings per household delivered due to the energy efficiency program

**Lifecycle net therm savings per household:** The total lifecycle net therms savings per household delivered due to the energy efficiency program

**Gross incentive per household:** The gross dollars that go directly to the customer/household.

**End use rebate per household:** The dollars refunded to the customer/household.

In addition, t-tests were performed to identify a statistically significant difference between the mean of the top quintile (20%) served and the bottom quintile (20%) served for each explanatory variable. Generally, the null hypothesis being tested is that there is no difference between the socio-demographic makeup of the population with the highest value for a given response variable and the lowest value of a given response variable. This question is posed to probe the equality of distribution of program service with respect to sociodemographic characteristics. This process is repeated after filtering for given program characteristics to probe the equality of distribution with respect to program characteristics. Particular program characteristics used are discussed further below (see section 3.4.2 for further explanation).

Spatial mapping was also used to assist in data visualization and exploration. For this task, QGIS was used throughout the analysis to visualize the data and findings through mapping (QGIS.org, 2020). This type of visualization is helpful for finding trends and clustering, as well as showcasing results and emphasizing the importance of spatial justice. In particular, the analysis explores the equity implications of the distribution of benefits by comparing the burden faced by the communities comprising the top and bottom quintiles of each considered variable.

### 3.3 Other Methods Explored

Additional methods of analysis were explored, including discriminant analysis and quantile regression. Discriminant analysis is similar to the logistic regression used by Reames et al. (2016) due to its use of categorical response variables. However, discriminant analysis differs in its ability to include more than two categories. The categorical variables were based on the program characteristics described in section 3.4.2 and results showed that each category could not be completely separated from each other by explanatory variables. As such, this analysis was not included in the result.

Two additional regression methods were explored in an attempt to alleviate certain problems with the distribution of response and explanatory variables. Quantile regression was explored to discern varying relationships to the response variable within a given demographic. Mixed effect linear modeling was also explored in an attempt to account for interaction effects across variables. A variety of variable selection methods were explored as well, in an attempt to isolate the most predictive explanatory variables. Specifically, ridge regression, lasso regression, and stepwise variable selection were explored. These additional analytical techniques resulted in ambiguous or unreasonable outcomes.

### 3.4 Data Sources

Multiple data sources were used to analyze the equitable distribution of energy efficiency programs across geographic and sociodemographic groups. The data broadly consists of two types: energy efficiency program claims data, which broadly define the response variables in this analysis and sociodemographic census data which broadly describe the explanatory variables.

Claims data were derived from a dataset provided by Lara Ettenson of the Natural Resources Defense Council (NRDC), which included California demand side management claims data from the California Energy Data and Reporting System (CEDARS) disaggregated by zip code, program ID, and year (2017-2019) (California Public Utilities Commission, 2020). These data are available through the online web portal for CEDARS, but not disaggregated by zip code. Further, certain program characteristics were missing from the provided claims data disaggregated by zip code, so program data from 2017-2019 were downloaded directly from CEDARS and were elaborated to create additional variables related to program characteristics.

Sociodemographic data were derived from the 2018 U.S. Census Bureau ACS 5-year average in zip code tabulation areas (ZCTAs) (U.S. Census Bureau, 2018). Lastly, information on the land area of the ZCTAs was sourced from R package Tigris (Walker, 2020).

The culmination of these sources ultimately led to two different datasets: an aggregated sum of all claims data at the zip code level and an aggregated sum of claims data at the zip code level filtered by program characteristics. Both of these claims-related datasets were then merged with the zip code level sociodemographic and land area data.

#### 3.4.1 Zip Code-Based Claims Data

After merging the zip code-disaggregated claims data with the supplementary CEDARS program information, the data were filtered based on programs for which the primary sector was residential. The claims data were then summed up by zip code. This data frame was then merged with the census data by zip code. Next, the land area information from TIGER was used to divide by total population (from census data) to create a population density variable.

#### 3.4.2 Program Data Collection

For each program that had claims in 2017-2019 and had the primary sector listed as residential, the program implementation plan was downloaded from CEDARS and reviewed to add supplementary program data into the analysis. This supplementary information was included by the addition of the following categorical variables: housing type, barriers to entry, and low-income targeted programs. The variable housing type was used to identify if a program serves single family housing, multifamily housing, or both. If the information was not found, the programs were assumed to serve both. Programs where participants must meet certain criteria to participate, such as owning a mobile home or living in a certain city or climate zone, were classified as having high barriers to entry. Programs were classified as having medium barriers to entry if they required criteria such as owning a smart thermostat, being able to troubleshoot technical issues, the requirement to contribute over 50% of project costs, completing building performance modeling, or being a high energy user. All other programs were classified as having low barriers to entry. Programs were assumed to target low-income customers, and classified as such, if they were administered by a regional energy network, targeted certain cities, were full direct install, or targeted mobile homes. Other program features, such as if a program is categorized as downstream, midstream, or direct install, were gathered directly from the CEDARS program data.

The supplementary program data was then used to filter the claims dataset and produce new datasets, identical in generation from the zip code-based claims data except for filtering by a particular program before summing up claims by zip code. Unique datasets from these program variables include low-income targeted programs, high monetary barriers to entry, downstream programs, and midstream programs. Additional datasets were generated filtering for claims only in 2017 and 2019.

#### 3.4.3 Data Processing

To prepare the data for the analysis, outliers having a Cook's Distance of more than 4 times the mean were removed. Cook’s distance is a technique used to identify influential outliers that negatively impact the reliability of the model. This is completed by running the regression model without the zip code and then observing the magnitude of change. The outliers, therefore, differ between the equations and depend on the response variables used. Next, to avoid overfitting the model, collinearity between explanatory variables was checked using Variance Inflation Factors (VIF). The variables ‘English as a second language’ and ‘unemployment rate’ had a VIF over 10 and were therefore removed.

#### 3.4.4 Data Limitations

The provided 2017-2019 claims data had zip code as the location granularity, which is quite a large aggregation area when considering the number of communities and neighborhoods included in a single zip code.

Further, the CEDARS data variables were limited. For example, required consumer dollar investment for program participation was not included, but program spending by the program administrator was included. The omission of customer-related costs prevents a real understanding of barriers to entry.

## 4. Results

### 4.1 Zip Code Level Analyses

#### 4.1.1 Summary Statistics

Summary statistics for the explanatory variables are listed in Table 1 below. The first column describes the explanatory variable and the second reports the average value across all zip codes in the analysis area. The third column is filtered for the zip codes that are in the bottom quintile for 4 or 5 of the response variables: participation rate, lifecycle net kilowatt-hour (kWh) savings per household, lifecycle net therm savings per household, gross incentive per household, end use rebate per household. Throughout this report, the term “underserved zip codes” refers to those zip codes in the bottom quintile for 4 or 5 of the considered response variables as in the summary statistics below, or the bottom quintile for the response variable being analyzed in a given test or regression.

Table 1. Summary statistics for explanatory variables across all zip codes.

|  |  |  |
| --- | --- | --- |
| Explanatory Variable | Average | Underserved |
| Percent white | 67.679 | 77.952 |
| Percent Hispanic | 33.16 | 27.195 |
| Percent Asian | 10.856 | 3.315 |
| Percent black | 4.45 | 2.831 |
| Percent Native American | 1.232 | 3.404 |
| Percent Hawaiian | 0.31 | 0.187 |
| Percent foreign born | 22.433 | 14.406 |
| Median age | 39.548 | 42.628 |
| Average household size | 2.888 | 2.655 |
| Percent old housing | 9.629 | 12.834 |
| Percent living in the same house as prior year | 86.624 | 86.172 |
| Percent renter | 41.822 | 42.254 |
| Percent 20 or more units in housing | 8.912 | 4.352 |
| Percent mobile home | 6.26 | 16.428 |
| Median household income | 74090.371 | 49904.035 |
| Percent with high gross rent as a percentage of income | 45.103 | 44.962 |
| Unemployment rate | 15.119 | 11.538 |
| Percent poverty | 14.81 | 20.207 |
| Percent highschool degree or higher | 83.515 | 81.581 |
| Percent with broadband internet | 81.928 | 71.18 |
| Percent disability | 11.985 | 15.797 |
| Percent parents working w/child | 61.055 | 56.149 |
| Percent no healthcare | 8.116 | 10.33 |
| Population density | 0.002 | 0.001 |

##### **4.1.2 Spatial Characterization of Data**

The following figures depict the spatial distribution of various CEDARS data by zip code. Since CEDARS data is reported by zip code, this information was combined with existing zip code shapefiles (i.e., the files used by GIS software to create shapes at the desired location along with associated data) to present it visually. The zip code shapefiles were developed by the U.S. Census Bureau using 2010 census data (U.S. Government,2019). Ultimately, five variables were standardized on a per household basis and presented spatially: gross incentive received, net kWh savings, net therm savings, total unit rebate across all claims, and the number of installed units.

Additional figures were created to visualize the distribution of benefits as a function of other environmental and programmatic features. These include: mapping net kWh and net therm savings by climate zone, and comparing the distribution of programs to urban areas. These figures can be found in the Appendix Section A3.

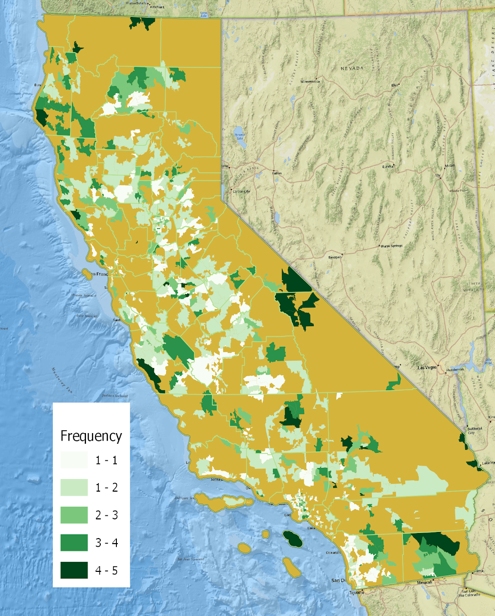


Figure 1. Frequency of zip codes in bottom quintile for response variables (participation rate, lifecycle net kilowatt-hour (kWh) savings per household, lifecycle net therm savings per household, gross incentive per household, and end user rebate per household).



Figure 2. Zip codes in the lowest and highest quintile for gross incentive per household for all claims (2017-2019).

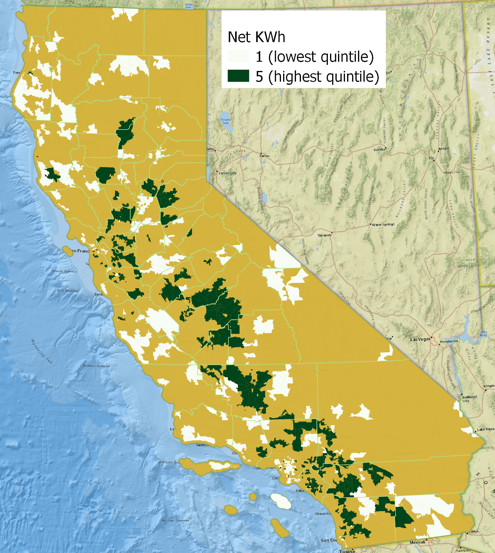


Figure 3. Zip codes in the lowest and highest quintile for total net kWh savings per household for all claims (2017-2019).

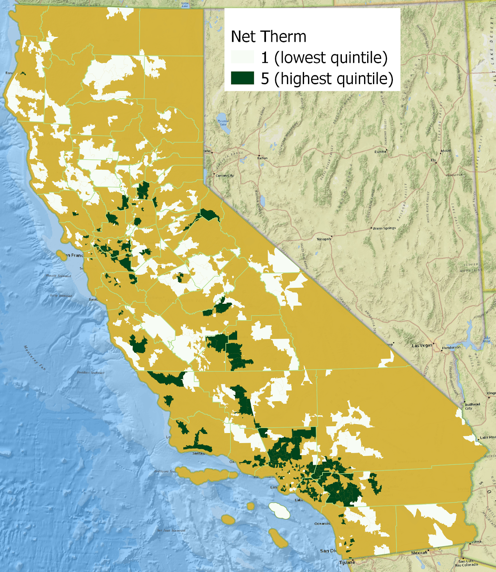


Figure 4. Zip codes in the lowest and highest quintile for total net therm savings per household for all claims (2017-2019).

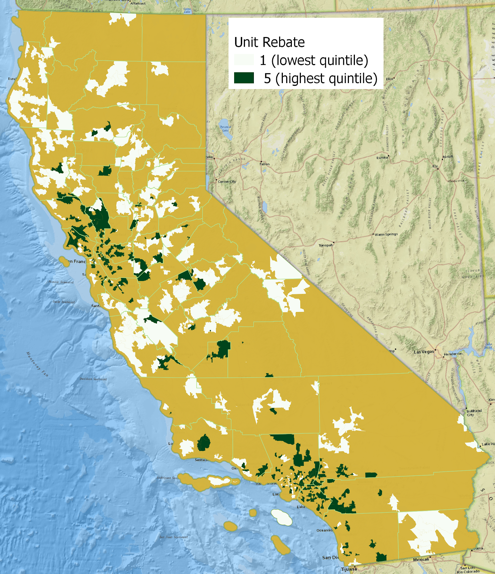


Figure 5. Zip codes in the lowest and highest quintile for the total unit rebate per household for all claims (2017-2019).



Figure 6. Zip codes in the lowest and highest quintile for the number of units installed per household for all claims (2017-2019).

To explore whether the distribution of benefits is equitable, the analysis compares the upper and lower quintiles from the examined CEDARS data (i.e., the 20% who receive the most and least benefits, respectively) to CalEnviroScreen 3.0 (CES) burden scores, which are shown in the Appendix (Figure A1). CalEnviroScreen is the state of California’s environmental health screening tool that reports the burden faced by communities across census tracts (OEHHA, 2018). The communities in the upper quartile of burden (i.e., the 75th percentile) are considered disadvantaged communities and are the center of many equity-based policies. The population-weighted average percentile burden score for the upper and lower quintiles of the five variables are calculated and compared. Under equitable benefit distribution, those with the higher burden scores should see the higher benefits from energy efficiency programs.

Table 2 reports a summary of the population-weighted average CalEnviroScreen percentile burden score for the upper and lower quintiles for each tracked variable. Note that, on average, communities who see the least benefits from energy efficiency programs tend to have higher burden scores than those who see the most benefits.

Table 2. Average population-weighted CES percentile burden score for the lower and upper quintiles of selected response variables.

|  |  |  |
| --- | --- | --- |
| Quintile | Average CES Percentile | |
| Unit Rebate | | |
| Lower quintile | | 52.535 |
| Upper quintile | | 47.141 |
| Number of units | | |
| Lower quintile | | 52.199 |
| Upper quintile | | 43.734 |
| Net kWh saved | | |
| Lower quintile | | 62.13 |
| Upper quintile | | 48.776 |
| Net therm saved | | |
| Lower quintile | | 52.624 |
| Upper quintile | | 50.016 |
| Gross incentive | | |
| Lower quintile | | 51.454 |
| Upper quintile | | 49.874 |

#### 4.1.3 Population Tests (T-tests)

T-test results for the bottom quintile, top quintile, and p-value of the mean difference between quintile populations are shown in Table 3. The difference between the two populations is considered significant if the p-value is less than 0.05. The prevailing trends are relatively constant throughout the different response variables. For underserved zip codes, there appears to be a higher percentage of poverty, old housing, population with a disability, population with no healthcare, population living in a mobile home, the population of Native Americans, and population renting. Additionally, for the underserved zip codes there seemed to be a lower population percentage of high school graduates, a lower population percentage with access to broadband internet, and a lower median household income. The response variable for the number of units per household in a given zip code showed additional relationships beyond the prevailing trend; namely, a higher percentage of the population comprised of Black or African Americans, Hispanics, those who are foreign-born, those living in housing with 20+ units, and those with a high Gross Rent as a Percentage of Income (GRAPI).

Table 3. T-test results for the bottom quintile, top quintile, and p-value of the mean difference between quintile populations (lq = lower quintile, uq = upper quintile).

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Variables | Num Units lq | Num Units uq | Num Units p-value | Net kWh lq | NetkWh uq | NetkWh p-value | Nettherms lq | Nettherms uq | Nettherms p-value | GrossIncentive lq | GrossIncentive uq | Gross Incentive p-value | User Rebate lq | User Rebate uq | UserRebate p-value |
| Percent Black or African American | 4.998 | 3.611 | 0.009 | 5.618 | 4.856 | 0.237 | 2.764 | 5.259 | 0 | 2.949 | 4.826 | 0 | 4.169 | 5.165 | 0.104 |
| Median Household Income | 65238 | 87276 | 0 | 56588 | 76542 | 0 | 54076 | 83571 | 0 | 588877 | 76017 | 0 | 59237 | 85012 | 0 |
| Unemployment Rate | 7.528 | 1.794 | 0 | 6.34 | 15.48 | 0.431 | 2.26 | 5.807 | 0 | 11.837 | 15.585 | 0 | 14.243 | 15.141 | 0.386 |
| Percent Hispanic | 39.375 | 26.761 | 0 | 34.321 | 37.979 | 0.084 | 30.165 | 35.671 | 0.006 | 28.057 | 38.627 | 0 | 33.03 | 31.586 | 0.476 |
| Percent Poverty | 18.283 | 12.454 | 0 | 19.807 | 14.801 | 0 | 19.577 | 12.848 | 0 | 18.106 | 14.98 | 0 | 18.538 | 12.43 | 0 |
| Percent Renter | 47.667 | 34.297 | 0 | 49.892 | 41.047 | 0 | 40.423 | 40.994 | 0.707 | 41.231 | 40.821 | 0.789 | 44.077 | 42.19 | 0.217 |
| Percent High School Degree or Higher | 80.18 | 87.699 | 0 | 79.961 | 83.132 | 0.009 | 80.569 | 85.493 | 0 | 82.728 | 82.826 | 0.936 | 81.651 | 86.025 | 0 |
| Percent Foreign Born | 23.397 | 20.608 | 0.015 | 21.206 | 23.463 | 0.04 | 15.882 | 25.59 | 0 | 15.552 | 23.422 | 0 | 18.619 | 24.596 | 0 |
| Percent with Broadband Internet Access | 79.87 | 85.062 | 0 | 75.424 | 84.198 | 0 | 73.677 | 86.363 | 0 | 75.271 | 83.917 | 0 | 77.604 | 85.32 | 0 |
| Percent Old Housing | 10.09 | 6.629 | 0 | 15.028 | 5.001 | 0 | 10.787 | 5.673 | 0 | 12.893 | 4.612 | 0 | 10.853 | 12.315 | 0.222 |
| Percent with Disabilities | 11.577 | 11.76 | 0.66 | 13.643 | 11.11 | 0 | 15.56 | 10.184 | 0 | 15.015 | 11.102 | 0 | 14.038 | 10.362 | 0 |
| Percent without Healthcare | 9.839 | 6.044 | 0 | 10.163 | 7.684 | 0 | 9.657 | 7.459 | 0 | 8.882 | 7.781 | 0.009 | 9.235 | 7.226 | 0 |
| Percent Mobile home | 6.944 | 4.63 | 0.005 | 9.18 | 5.271 | 0 | 13.691 | 3.859 | 0 | 11.832 | 5.439 | 0 | 8.89 | 3.496 | 0 |
| Percent Native American | 1.6 | 1.215 | 0.329 | 2.033 | 0.862 | 0 | 3.061 | 0.66 | 0 | 2.453 | 0.86 | 0 | 2.054 | 0.792 | 0.001 |
| High GRAPI | 47.426 | 44.036 | 0.004 | 45.905 | 45.671 | 0.839 | 45.466 | 46.237 | 0.551 | 45.423 | 45.958 | 0.681 | 45.525 | 44.327 | 0.323 |

##### 

##### 4.1.4 Multivariate Regression Results

Results from the multivariate regression analysis in table X indicate underserved zip codes (i.e., those in the lowest quintile) have an older population, a higher percentage of old housing, a higher percentage of mobile homes, and a higher percentage of residents living in the same house as the year prior. The equity of energy efficiency programs serving Asian populations is slightly more difficult to conclude due to zip codes with a higher percent Asian population having both a high participation rate and a lower net kWh savings per household.

Identifying the correct metric to determine a population as being served or underserved is difficult. The difficulty of defining it is made more evident by the differing results for each response variable. For example, if net kWh savings per household is the identifying factor of serving a household, zip codes with a higher percentage of the Asian population, a higher percentage of old housing, and a higher percentage of residents living in the same house as the year prior are underserved. If the participation rate is the identifying factor of serving a household, an older population, and a higher percentage of old housing are underserved. Despite the differences between the response variables, it is evident energy efficiency programs are not serving old housing to the same extent as newer housing, due to its statistical significance in three of the response variables analyzed.

Table 4. Multivariate regression results that have statistical significance for at least one of the response or explanatory results. The coefficient is listed first, followed by the standard errors in parentheses. \*, \*\*, \*\*\* indicates significance at the 90%, 95%, and 99% level, respectively. The table entries shaded in green or orange indicate coefficients with statistical significance. The green represents a positive correlation, and the orange represents a negative correlation. The full multivariate regression results are provided in Appendix Section A2.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Participation rate | Net kWh savings per household | Gross incentive per household | End use rebate per household |
| % white | 2.152 \*\*\* (0.000535) | 125 \*\*\* (0.000331) | 7.119 \*\*\* (0.000449) | -0.2171 (0.1821) |
| % black | 0.2731 (0.620342) | 80.95 \*\* (0.008963) | 8.811 \*\*\* (0.00000119) | -0.04035 (0.7826) |
| % Asian | 2.189 \*\* (0.001157) | -82.49 \* (0.028487) | 2.056 (0.350484) | 0.06219 (0.72716) |
| % Hispanic | -0.3436 (0.602413) | 33.06  (0.36826) | 6.285 \*\* (0.003561) | 0.5353 \*\* (0.00203) |
| Median age of population | -0.1231\* (0.014765) | -2.874  (0.306278) | -0.508 \*\* (0.002302) | 0.003539 (0.789) |
| % with internet | 0.02513 (0.473611) | 8.284 \*\*\* (0.0000241) | 0.4143 \*\*\* (0.00031) | 0.004617 (0.61547) |
| % old housing | -0.04991 \* (0.02386) | -4.698 \*\*\* (0.000146) | -0.3694 \*\*\* (0.000000404) | 0.004873 (0.412) |
| % same house as prior year | 0.04515  (0.22603) | -4.976 \*  (0.017418) | -0.2896 \* (0.021243) | -0.005834 (0.55633) |
| % working parents | 0.03066 \* (0.011287) | -0.105  (0.875798) | 0.00999 (0.800762) | -0.004306 (0.17807) |
| % mobile homes | -0.09792 \*\*\* (0.000316) | 0.1441  (0.923818) | -0.005763 (0.948172) | 0.00413 (0.56142) |
| % buildings with 20+ units | -0.001937 (0.941099) | 0.5032  (0.741887) | 0.2243 \*\* (0.009345) | 0.03371 \*\*\* (0.0000835) |
| R-squared | 0.4379 | 0.229 | 0.469 | 0.240 |
| Adjusted R-squared | 0.4287 | 0.217 | 0.461 | 0.228= |

### 4.2 Program-Filtered Analyses

Two supplementary characteristics were used in this filtering procedure: an indicator for low-income targeted programs and an indicator for high barriers to entry. Filtering for programs targeting low-income customers results in some erasure of underserved groups, most notably in Net kWh and Gross Incentive. Alternatively, when filtering for programs deemed to have high monetary barriers to entry, there is only one response variable that has anomalous underserved populations: Net kWh. There appears to be a larger population percentage of Black or African Americans and a population with high GRAPI in the underserved quintile. Though it should be noted, the difference in population means among response variables for median household income is anonymously negligible for this filter.[[1]](#footnote-1)

Two analyses were performed to look at the resulting population distribution when filtering for midstream programs and downstream programs. When filtered for programs designated as “Midstream,” there appear to be more instances of systematic under-service of particular sociodemographic characteristics than the 3-year average of all claims data. The distribution of service for programs designated as “Downstream” broadly adheres to the 3-year average.

Though not an explicit program characteristic, year was used as a filter for the claims data to approximate trends over time. In 2017, the relationship between sociodemographic groups and service is similar to the aforementioned 3-year average results where there is a discernible trend among the underserved quintiles. The notable exception is the increased percentage population of Black or African Americans for the underserved quintile in Net kWh savings/household. When filtering for claims in 2019, the trend in unserved zip codes erodes slightly. One particularly interesting change from both 2017 and the 3-year average is the relationship between Net kWh savings/household and the median household income.

## 5. Discussion and Conclusions

### 5.1 Spatial Distribution

One outcome of interest from spatial modeling arises from comparing the distribution of energy efficiency program benefits to CalEnviroScreen scores. Overlay analysis of these spatial data revealed that, on average, communities who saw the least benefits from energy efficiency programs had (sometimes negligibly) higher burden scores than those who saw the most benefits from these programs. At the very least, it appears that the distribution of benefits is nearly equal across all communities (as defined by CalEnviroScreen designations). This is substantiated by the fact that the average CES percentiles of the analyzed quintiles were near 50, meaning the programs reached the average population nearly uniformly. As equitable distribution would provide more benefits to those who face higher burdens, this *lack of difference* in burden between the upper and lower quintiles shows that current energy efficiency programs may be underperforming.

However, there are certainly explanations for why the lack of difference currently exists that complicate conclusion making. For example, the average percentile burden score of the quintile that saw the greatest net therms saved is 52. This means that, on average, a community with a very high burden score and one with a very low burden score saw the same amount of net therms saved. However, it is likely that these two communities have very different therm usage, as the high-burden community (which indicates, among other factors, low socioeconomic status) is likely to own a smaller home or be more frugal with natural gas use. Therefore, the number of therms reduced may constitute a larger proportion of the high-burden community’s total therm usage than that of a low-burden community. Thus, while on the surface there is no difference in net benefits gained by these two communities, the relative benefit under this scenario is larger for the high-burden community. A lack of difference in percentile burden score between the quintiles who achieved the fewest and the most net therm savings can, again, be explained by a difference in net therm usage across communities. A similar hypothesis can be made regarding electricity use, where the difference in average percentile between the upper and lower quintile is more pronounced. This suggests that the energy efficiency programs are functioning equitably. On the other hand, there may be no difference in the number of units installed across communities because the price of entry is still too high for low-income communities. Ultimately, these findings require more conclusive evaluation to validate the hypothetical scenarios discussed here.

### 5.2 What the analyses suggest about underserved populations

Overall, the statistical analyses by zip code demonstrate that populations underserved by energy efficiency programs are within zip codes that have older housing, lower income residents, a higher renting population, and a higher population of Native Americans. Results differ depending on the response variable and the method of analysis. A host of response variables were used in order to capture participation, energy savings, and dollar savings, and demonstrate the difficulty of defining what “being served” by a program means, as well as demonstrating the heterogeneity between the way people are benefiting or being reached by the programs.

The bottom quintile of each response variable was identified to indicate the underserved zip codes, although the results between the multivariate regression and the quintile t-test analyses, while keeping the response variables the same, also differed. The multivariate regression, for example, did not identify many socioeconomic categories as an indicator of populations being served (i.e., residents under the poverty line, older housing, a higher population with a disability, a higher population with no healthcare, a higher population of Native Americans, and a higher renting population); however, the t-test showed statistical significance between the top and bottom quintiles. These findings demonstrate that these variables may not be a predictor of program benefits, but that there is still not an equal distribution of benefits among energy users in relation to their socioeconomic characteristics.

### 5.3 The effect program type on highest and lowest quintiles

The analysis of claims data, when filtered by program characteristics, yielded some interesting results. If there is a desire to increase service to certain socio-demographic groups, quantifying the level of service provided by the different tools at program administrators’ disposal appears to be important. Claims data filtered by midstream programs and downstream programs showed more instances of sociodemographic groups with a higher population average in the lowest quintile for Black or African American, Hispanic, and high GRAPI. A midstream designation refers to a program targeting contractors and equipment distributors. This finding may echo findings by other researchers, such as those in Reames et al. (2018) that identify a lack of access to energy efficiency products for certain demographic groups due to the types of distributors or contractors that are targeted by these midstream programs. Given that the distribution of service for downstream programs had fewer instances of sociodemographic groups underperforming, midstream programs may have a propensity for less equitable distribution than downstream programs. Downstream programs that explicitly target low-income customers may be capable of delivering improved equity of energy efficiency services. In addition, those programs that have a high monetary barriers-to-entry may disproportionately benefit wealthier customers, and thus quantifying the effect of these programs on the distribution of energy efficiency program benefits among different populations is also important.

### 5.4 Analytical limitations resulting from zip code level aggregation of claims data

The analysis of energy efficiency programs and their respective distribution is ultimately hindered by the spatial aggregation at the zip code level. What this analysis can provide is a reflection of how these programs are distributed across California with low resolution. The goal of this analysis was to quantify how these programs are distributed across sociodemographic groups; however, because of data aggregation at the zip code level, the socio-demographic makeup of those being served is grossly represented by demographics at that scale. As a consequence, it is impossible to know how the claims are distributed across socio-demographic groups within a zip code, or even how they are distributed at more refined spatial scales. Ultimately, this could hide much more conclusive evidence of which populations may be underserved by energy efficiency programs.

Further, it is hard to discern differences between socio-demographic groups that may result from interactions with other variables. For example, if there is an interaction effect between a race characteristic and household income, this interaction is harder to identify because both of those variables are aggregated at zip code level, obscuring the individual or neighborhood level differences. Ideally, to perform a true socio-demographic gap analysis, claims data are needed at the household level. With household level claims data, the service can explicitly be attached to a household with a set of socio-demographic characteristics. These data could still be aggregated at zip code level or other spatial scales that sufficiently anonymize potentially identifiable information, as long as each claim includes sociodemographic information. An alternative, aggregating at the programmatic level might also improve future analyses and is described in section 5.5.

## 5.5 Limitations and suggestions for CEDARS program data

The CEDARS database provided the claims data and program information relevant to this project and is the public dataset available for evaluation. The claims data details information about program administrator spending and resulting savings, but does not capture any information about the program participants. To better analyze the types of participants in residential energy efficiency programs, it would be helpful to have program administrators capture high level sociodemographic information and home energy usage information about program participants and report to CEDARS. In addition, reporting required participant investment to participate would help to better understand the monetary barriers to entry for different programs.

The program information data from CEDARS could be more informative for this analysis if it had key indicators for populations the program is intended to target. For example, noting if a program is intended to target lower income participants or participants in older housing would be useful. The analysis in this report developed criteria to estimate which programs were intended to target lower income participants, but having program administrators who know the details of the program report this data would be an improvement. Similarly, reporting if the program targets single family, multifamily, or both would be useful.

The data dictionary for CEDARS claims could be better developed to more clearly describe what each category of data represents. For example, “NumUnits'' was the best information reported in the claims to understand how many “units” of a program were delivered in a zip code but depending on the program, it is unclear if this is a piece of equipment or a whole home upgrade. The data dictionary reports that this metric should be used in conjunction with “NormUnit”, but this detail was not provided in the claims. Further, the distinction between rebates and incentives delivered was not clear from the data dictionary.

## Works Cited

Ambrose, A.; Baker, W.; Batty, E.; Hawkins, A. (2019). Reaching the ‘Hardest to Reach’ with energy advice: final report. *Sheffield Hallam University.* [DOI: 10.7190/cresr.2019.8286642862]

Brown, M.; Soni, A.; Lapsa, M.; Southworth, K. (2020). Low-Income Energy Affordability: Conclusions from a Literature Review. *Oak Ridge National Laboratory*.

California Public Utilities Commission (2020) CEDARS: California Energy Data and Reporting System. https://cedars.sound-data.com/

Chester, L. (2014). Energy Impoverishment: Addressing Capitalism’s New Driver of Inequality. *Journal of Economic Issues,* 48:2, 395-404.<https://doi.org/10.2753/JEI0021-3624480213>

Dolšak, J., Hrovatin, N. & Zorić, J. Factors impacting energy-efficient retrofits in the residential sector: The effectiveness of the Slovenian subsidy program. *Energy Build.* **229**, (2020).

Elnakat, A.; Gomez, J.;Booth, N. (2016). A zip code study of socioeconomic, demographic, and household gendered influence on the residential energy sector. *Energy Reports,* 2:21-27.<http://dx.doi.org/10.1016/j.egyr.2016.01.003>

Hernández, D., Jiang, Y., Carrión, D., Phillips, D. & Aratani, Y. Housing hardship and energy insecurity among native-born and immigrant low-income families with children in the United States. *J. Child. Poverty* **22**, 77–92 (2016).

Forrester, S.; Reames, T. (2020). Understanding the residential energy efficiency financing coverage gap and market potential. *Applied Energy,* 260(114307): 1-12.<https://doi.org/10.1016/j.apenergy.2019.114307>

Leventis, G.; Fadrhonc, E.; Kramer, C.; Goldman, C. (2016). Current Practices in Efficiency Financing: An Overview for State and Local Governments. *Berkeley National Laboratory*.<https://emp.lbl.gov/sites/all/files/lbnl-1006406.pdf>

OEHHA (2018). SB 535 Disadvantaged Communities using CalEnviroScreen 3.0 (updated June 2018). Available online: https://oehha.ca.gov/calenviroscreen/sb535. Accessed 15 Dec, 2020.

QGIS.org, 2020. QGIS Geographic Information System. QGIS Association. [http://www.qgis.org](https://www.qgis.org/)

Reames, T. (2016a). A community-based approach to low-income residential energy efficiency participation barriers. *Local Environment.* 21:12, 1449-1466.<https://doi.org/10.1080/13549839.2015.1136995>

Reames, T. (2016b). Targeting energy justice: Exploring spatial, racial/ethnic and socioeconomic disparities in urban residential heating energy efficiency. *Energy Policy,* 97:549-558.<http://dx.doi.org/10.1016/j.enpol.2016.07.048>

Reames, T.; Reiner, M.; Ben Stacey, M. (2018). An incandescent truth: Disparities in energy-efficient lighting availability and prices in an urban U.S. county. *Applied Energy,* 218:95-103.<https://doi.org/10.1016/j.apenergy.2018.02.143>

Reames, T.; Stacey, B.; Zimmerman, M. (2019). A Multi-State Analysis of Equity in Utility-Sponsored Energy Efficiency Investments for Residential Electric Customers. *Poverty Solutions.* https://poverty.umich.edu/files/2019/05/Energy\_efficiency.pdf

Sokolova, A.; Aksanli, B. (2019). Demographical Energy Usage Analysis of Residential Buildings. *Journal of Energy Resources Technology,* 141. [DOI: 10.1115/1.4042451]

U.S. Energy Information Administration (2020). California State Energy Profile. Available at: <https://www.eia.gov/state/?sid=CA#tabs-2>. Last accessed 12/13/2020.

United States Government (2020). TIGER/Line Shapefile, 2019, 2010 nation, U.S., 2010 Census 5-Digit ZIP Code Tabulation Area (ZCTA5) National. Available at: <https://catalog.data.gov/dataset/tiger-line-shapefile-2019-2010-nation-u-s-2010-census-5-digit-zip-code-tabulation-area-zcta5-na>. Last accessed 12/15/2020.

Walker, K. (2020). tigris: Load Census TIGER/Line Shapefiles. R package version 1.0. <https://CRAN.R-project.org/package=tigris>

Xu, X.; Chen C. (2019). Energy Efficiency and Energy Justice for U.S. Low-Income Households: An Analysis of Multifaceted Challenges and Potential. *Energy Policy,* 128:763-774.<https://doi.org/10.1016/j.enpol.2019.01.020>

Zimring, M.; Borgeson, M.; Hoffman, I.; Goldman, C.; Stuart, E.; Todd, A.; Billingsley, M. (2011). Delivering Energy Efficiency to Middle Income Single Family Households. *Lawrence Berkeley National Laboratory*.<https://www7.eere.energy.gov/seeaction/system/files/documents/middle_income.pdf>

## Appendix

### Section A1. Explanatory Variables

Explanatory variables: Percent white, Percent Black, Average household size, Median household income, Median age of residents, Percent Asian, Percent Hispanic, Percent poverty, Percent renter, Percent with high school degree, Percent foreign-born, Percent with internet, Percent old housing, Percent same house as prior year, Percent with disability, Percent of working parents with child, Percent with no healthcare, Percent mobile homes, Percent of buildings with 20+ units, Percent Native American, Percent Hawaiian, High GRAPI, Population density, Unemployment rate

### Section A2. Multivariate regression results

Table A1. Complete multivariate regression results. The coefficient is listed first, followed by the standard errors in parentheses. \*, \*\*, \*\*\* indicates significance at the 90%, 95%, and 99% level, respectively.

|  | Participation rate | Net kWh savings per household | Net therm savings per household | Gross incentive per household | End use rebate per household |
| --- | --- | --- | --- | --- | --- |
| Percent White | 2.152 \*\*\* (0.000535) | 125 \*\*\* (0.000331) | -6.199. (0.0532) | 7.119 \*\*\* (0.000449) | -0.2171 (0.1821) |
| Percent Black | 0.2731 (0.620342) | 80.95 \*\* (0.008963) | 3.574  (0.2106) | 8.811 \*\*\* (0.00000119) | -0.04035 (0.7826) |
| Average household size | 0.4916 (0.482962) | -13.35 (0.732054) | 7.012  (0.0548) | 3.897  (0.098309) | -0.02759 (0.8822) |
| Median household income | -0.000001438  (0.91089) | -0.00007844 (0.912583) | 0.00004345 (0.5154) | -0.000001158 (0.978044) | 0.000003892 (0.25702) |
| Median age of residents | -0.1231\* (0.014765) | -2.874 (0.306278) | 0.05783 (0.8254) | -0.508 \*\* (0.002302) | 0.003539 (0.789) |
| Percent Asian | 2.189 \*\* (0.001157) | -82.49 \* (0.028487) | 2.148  (0.5388) | 2.056 (0.350484) | 0.06219 (0.72716) |
| Percent Hispanic | -0.3436 (0.602413) | 33.06 (0.36826) | -3.617 (0.2905) | 6.285 \*\* (0.003561) | 0.5353 \*\* (0.00203) |
| Percent poverty | 0.01797 (0.630579) | 2.395 (0.251319) | -0.128 (0.5081) | 0.134 (0.273907) | 0.002093 (0.83204) |
| Percent renter | -0.02103 (0.331377) | 0.4262 (0.723015) | -0.03557 (0.7518) | -0.09574 (0.175868) | -0.009369 (0.10486) |
| Percent with high school degree | 0.02201 (0.531419) | -0.3613 (0.85323) | -0.07208 (0.6933) | 0.1266 (0.271395) | 0.01039 (0.26547) |
| Percent foreign-born | -0.01617 (0.626775) | 1.676 (0.363874) | -0.04097 (0.8125) | 0.06735 (0.535766) | 0.0003244 (0.97104) |
| Percent with internet | 0.02513 (0.473611) | 8.284 \*\*\* (0.0000241) | 0.2067 (0.2564) | 0.4143 \*\*\* (0.00031) | 0.004617 (0.61547) |
| Percent old housing | -0.04991 \* (0.02386) | -4.698 \*\*\* (0.000146) | 0.0602 (0.6051) | -0.3694 \*\*\* (0.000000404) | 0.004873 (0.412) |
| Percent same house as prior year | 0.04515 (0.22603) | -4.976 \* (0.017418) | -0.2875 (0.1379) | -0.2896 \* (0.021243) | -0.005834 (0.55633) |
| Percent with disability | 0.0822 (0.176799) | -3.979 (0.238157) | -0.004135 (0.9895) | 0.0293 (0.884427) | -0.006984 (0.66211) |
| Percent of working parents with child | 0.03066 \* (0.011287) | -0.105 (0.875798) | 0.03412 (0.5869) | 0.00999 (0.800762) | -0.004306 (0.17807) |
| Percent with no healthcare | -0.07315 (0.250272) | 4.167 (0.239377) | -0.4635 (0.1581) | 0.05238 (0.800493) | -0.01984 (0.23424) |
| Percent mobile homes | -0.09792 \*\*\* (0.000316) | 0.1441 (0.923818) | 0.1193 (0.3973) | -0.005763 (0.948172) | 0.00413 (0.56142) |
| Percent of buildings with 20+ units | -0.001937 (0.941099) | 0.5032 (0.741887) | 0.2317 (0.1094) | 0.2243 \*\* (0.009345) | 0.03371 \*\*\* (0.0000835) |
| Percent Native American | 0.02879 (0.960447) | -11.7 (0.716146) | 3.431  (0.2544) | -0.8114 (0.668263) | -0.2527. (0.09675) |
| Percent Hawaiian | -0.1044 (0.835705) | 42.25 (0.134901) | -3.295 (0.2079) | -0.9678 (0.556129) | 0.141 (0.29161) |
| High GRAPI | 0.01412 (0.417924) | 0.4305 (0.65777) | 0.07704 (0.3962) | 0.06053 (0.288636) | 0.004358 (0.34573) |
| Population density | -266.8. (0.064846) | 1394 (0.863766) | -567  (0.4597) | -471.7 (0.322804) | 76.92. (0.06605) |
| R-squared | 0.4379 | 0.229 | 0.06956 | 0.469 | 0.240 |
| Adjusted R-squared | 0.4287 | 0.217 | 0.05445 | 0.416 | 0.228 |

### Section A3. Additional spatial mapping figures

Map

Description automatically generated

Figure A1. CalEnviroScreen percentile burden scores. Those with scores in the upper quartile (dark red) are considered disadvantaged communities.

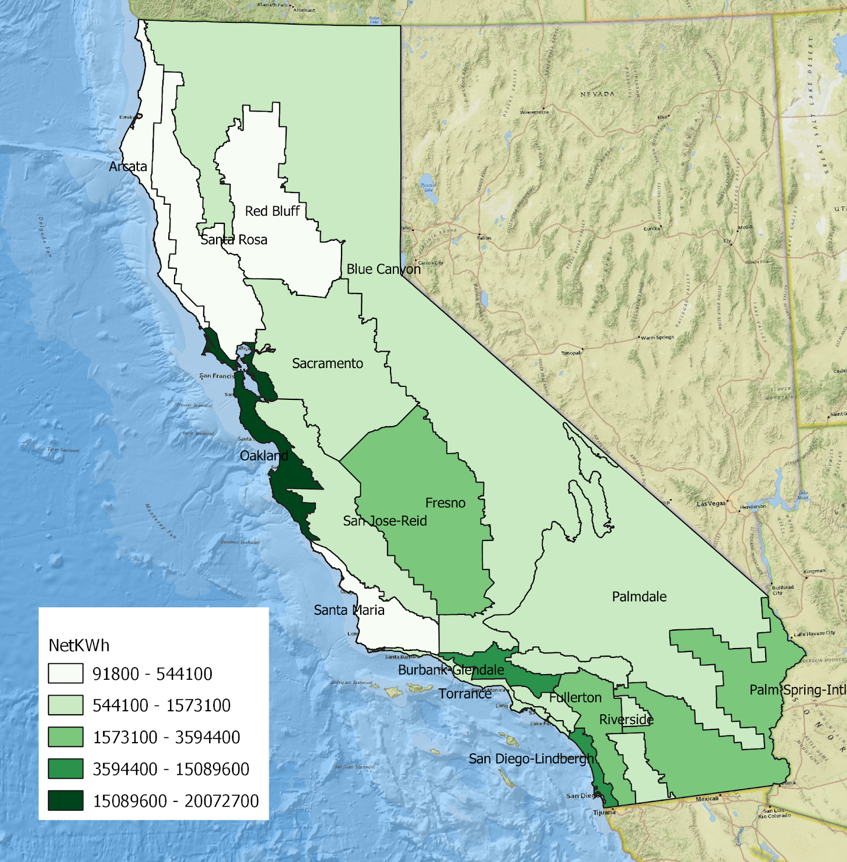


Figure A2. The mean total net KWh savings of energy efficiency programs across different climate zones.



Figure A3. The mean total net therm savings of energy efficiency programs across different climate zones.

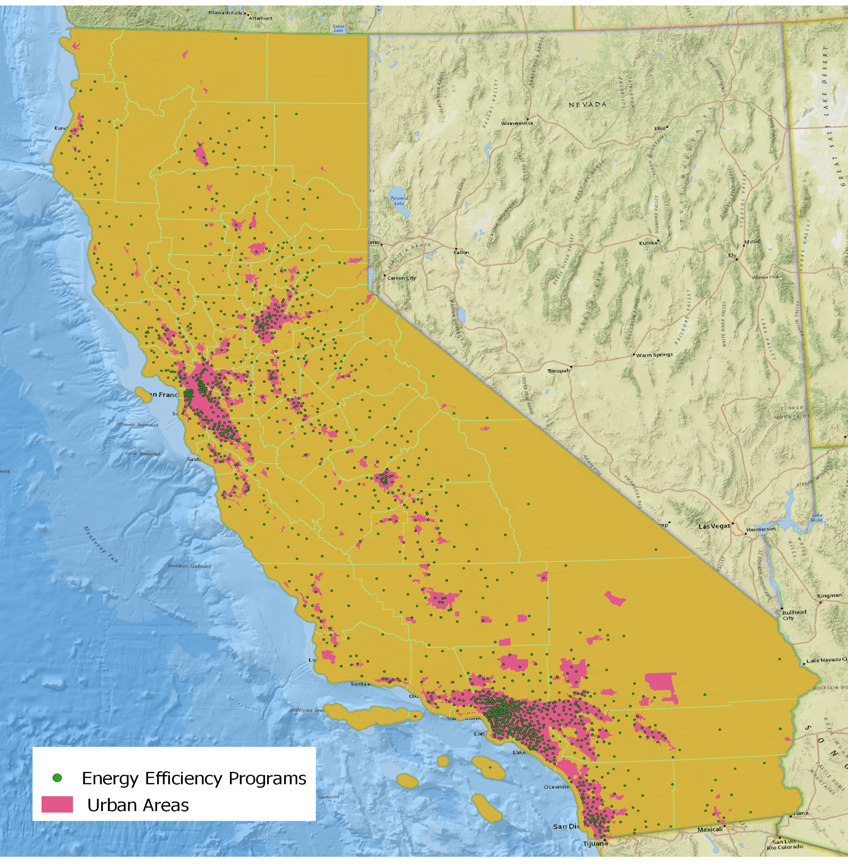


Figure A4. The distribution of the energy efficiency programs compared to the location of urban areas in California.

1. The content of these tables are all available upon request (file name: ttest\_results\_bucket3\*.csv (2017,2019, midstream, downstream, BTEM, lowincome) [↑](#footnote-ref-1)