Are Residential Energy Efficiency Programs Effective? An Empirical Analysis in Southern California

Yating Chuang
National Taipei University
yating@gm.ntpu.edu.tw

Magali Delmas
UCLA
delmas@ucla.edu

We evaluate 2010-2015 multiple residential energy efficiency subsidy programs using more than 11 million households' electricity billing records in Southern California. We find that these programs reduce overall electricity usage by 4 percent. However, there are significant differences in their effectiveness. Pool pumps and refrigeration programs generate the largest savings (12 percent, and 6 percent, respectively). Lighting and HVAC retrofits programs generate less than 1 percent savings. We also find electricity consumption increases for dishwashers and clothes washers upgrades, building shell, and whole house retrofits. Program impact varies during certain time of the year and for certain building types. (JEL L68, L94, Q41)

Keywords: Energy efficiency, electricity consumption, residential homes, residential real estate, program evaluation, energy policy

We thank project manager Felicia Federico for her essential contribution to the project management and her comments on the paper. We thank Dan Cheng and Hannah Gustafson for their assistance on the data request and geocoding of the building data. We thank the California Energy Commission (CEC) (GFO 15-312), the UCLA Ziman Center for Real Estate's Rosalinde and Arthur Gilbert Program in Real Estate, Finance and Urban Economics for generous funding. For helpful comments, we thank Eric Fournier, Patricia Yañez-Pagans, Omar Asensio, Alan Barreca, Erik Porse, Jinghui Lim, John Chung-En Liu, participants at the UCLA IoES research seminar, 19th Annual CU Environmental and Resource Economics Workshop, USC 10th International Workshop on Empirical Methods in Energy Economics, the Claremont McKenna College research seminar, and project stakeholders at the Advanced Energy Community (AEC) project team, including energy experts from the Energy Coalition and policymakers from the LA county.

I. Introduction

Energy efficiency (EE) is one of the main policy tools for addressing climate change. Former U.S. Secretary of Energy Steven Chu once said "If I were emperor of the world, I would put the pedal to the floor on energy efficiency and conservation for the next decade" (The Guardian 2009). EE subsidies are politically attractive because of their ability to both reduce energy usage and save money for consumers and for governments alike. As a result, governments around the world are developing policies to encourage energy efficiency. California, for example, passed a law mandating the ambitious target of reducing its carbon emissions by 40 percent below the 1990 level by 2030. One critical pathway identified to achieve this goal is to subsidize energy efficiency upgrades.[1] Indeed, the state of California spends about $1 billion dollars annually on residential energy efficiency upgrade programs.[2]

There is, however, limited empirical evidence about the effectiveness of these programs. Most claims regarding savings from energy efficiency upgrades, such as the famous McKinsey's cost curve, are based on engineering modeling projections (McKinsey and Company, 2009). Such projections usually ignore the behavioral aspects of energy consumption. For example, Chen et al. (2015) showed that households differ significantly in how they use the same appliances in similar apartments. Ignoring these behavioral differences in consumption might lead to an erroneous estimate of the energy savings. In fact, the recent empirical evidence of EE programs using experimental or quasi-experimental design suggests that seldom do these programs deliver the savings predicted by engineering estimates (Davis et al. 2014; Fowlie et al. 2015; Zivin and Novan 2016; Allcott and Greenstone 2017; Liang et al. 2017).

Scholars have argued that there is a need for more rigorous empirical research to identify heterogeneous program impacts based on program design, household and building characteristics (Allcott and Greenstone 2012). However, the difficulty in accessing energy consumption data,

[1] The CPUC's 2016 report states: "When California makes plans for new energy resources, energy efficiency is the state's first priority." California has been ranked among the top 3 states in the past five years by the American Council for an Energy Efficiency Economy (ACEEE) in terms of its resources devoted to energy efficiency.

[2] During the program cycle in our research, the CPUC authorized $3.1 billion ratepayer's money in funding 2010-2012 EE program. Investor-owned utilities (IOUs) spent most of the money—$2.6 billion of the total program money were spent by IOUs.

and the low take-up rates of EE programs hamper such empirical analyses. For example, Fowlie et al. (2015) conducted the nation's largest randomized experiment to encourage households to adopt a free energy retrofit program in Michigan. The take-up rate was 6% despite ample resources spent for recruitment. The low take-up of EE programs makes program evaluation difficult because of potential selection bias.[3] Even in some of the best-designed randomized control trials (RCT),[4] the low take-up rate of EE programs poses a significant barrier to instrument a randomization design as studies need large sample sizes to achieve enough statistical power.[5] This problem also hinders the external validity of the analysis.

Against this background, we have compiled a large dataset to evaluate the effectiveness of residential EE upgrade programs. We have access to meter-based monthly electricity consumption data for approximately 11 million households' in Southern California Edison (SCE) territory from 2010 to 2014. The data is from the LA Energy Atlas, a relational database that enhances understanding of energy usage across LA County (Pincetl and LA Energy Atlas Development Team, 2015). We also have information on all SCE residential EE upgrade programs during that period with all financial incentive records claimed by SCE customers to improve their home energy efficiency. The data allows us to address the following research questions: (1) How effective are residential EE programs (appliance and equipment upgrades) at saving energy? Does program effectiveness vary across income groups and building characteristics? (2) What is the difference between engineering estimates of potential energy savings and actual measured energy savings?

Our study makes two contributions. First, we refine our understanding of the general effectiveness of EE programs by analyzing multiple programs. To date, only a few studies use large-scale micro data in the U.S. Two prominent studies focus on residential energy efficiency

3 The low EE take-up can be because of imperfect information, split incentives between renters and homeowners, and some other behavioral bias, such as inattention. (Allcott and Greenstone 2012; Gillingham et al. 2012)

4 There are two large-scale RCT research cases in the literature with 30,000 observations in Fowlie et al. (2015) and 100,000 observations in Allcott and Greenstone (2017).

5 Sample size needed is inversely proportional to the square of p, where p is the difference in the proportion of the treatment group that takes up a program relative to the control group. So p=0.05 (0% of the control group, and 5% of the treatment group), researchers need 400 (=1/(0.05)^2) times the sample as with 100% compliance. We can do a simple exercise using our data information as a reference. Say average electricity usage is 600 kWh per month for non EE participants (control group), standard error is about 700 kWh per month. To detect 10% decrease in electricity consumption as treatment effect with 100% take-up, we need sample size of 2861; however, with only 5% take-up, we need sample size of 1,144,142.

programs and low-income weatherization programs in Wisconsin and Michigan (Fowlie et al., 2015; Allcott and Greenstone, 2017). Both studies find that energy efficiency programs are not cost-effective. However, because these studies focus on a single energy efficiency program (e.g. weatherization),[6] and/or a single demographic (e.g. low income), it is unclear whether the energy efficiency gap—defined by the difference between optimal and actual energy use—occurs because of the program or because of household demographics. To our knowledge, we provide the first study comparing numerous EE programs with large-scale micro data (e.g. more than 11 million households' utility billing records). Our comprehensive coverage of program data allows us not only to compare programs, but also to control for overlapping program impacts (owing to the fact that EE participants may apply to multiple programs) and avoid over-estimation that may result from these overlaps. With this unique large-scale data we can precisely compare impacts across program types (e.g. refrigerator, lighting, HVAC, etc.) with large statistical power. We can also investigate differences in program impact according to different incentive schemes, such as free giveaways and cash rebates.

Second, we are able to differentiate program impacts based on building characteristics, such as vintage, and square footage. This differs from most previous studies that link account level electricity information solely to census socio-demographics, rather than including building level information. Observing heterogeneous program effects is critical to improving funding allocation and program design.

To evaluate program impacts, we follow the identification strategy of the most rigorous research in EE program evaluation to date (Davis et al. 2014; Fowlie et al. 2015).[7] We rely on a difference-in-difference strategy, a rich number of matching covariates, and various fixed effects (household-month and time) to address self-selection bias in program participation since residential EE programs are voluntary. First, we construct a set of households that are similar to our program participants but have never participated in any of the programs. To achieve this, we use the Mahalanobis matching method on a pool of more than 10 million households based on

6 Allcott and Greenstone (2017) focus on two energy efficiency programs in Wisconsin, which are part of the national Better Building Neighborhood Program (BBN). The two programs were targeted at residential building retrofit. First, the energy consultant will give a free audit, and then recommend many parts of the needed retrofit to the building, such as attic insulation, air sealing, vaulted ceiling insulation, etc.

7 We do not have a valid instrumental variable or a credible regression discontinuity (RD) design as programs are available to all SCE customers in the SCE service territory.

building characteristics geocoded from the assessors' database and census blockgroup data.[8] Second, we use panel regression models to estimate the average program effect after matching. We exploit the monthly variation of program participation, and use household-month and time fixed effects to rule out time-invariant unobservable factors that might confound program impacts. Our identification is based on panel regressions comparing monthly electricity usage over time for households participating in EE programs to those who have never participated in the programs.

Our results show that EE programs reduce overall electricity usage by 4 percent. However, the energy savings vary significantly by program type. While pool pumps and refrigeration programs are associated with significant energy savings (12-13% and 6%, respectively), HVAC retrofits generate insignificant overall savings, and programs targeted at other appliances (i.e. clothes washers and dishwashers) and building shell, are associated with increases in overall consumption.

In addition, we find that energy savings are considerably inconsistent with the engineering estimates computed by Southern California Edison, the investor- owned utility for most of Southern California responsible for the EE incentive programs. For example, lighting programs achieve only 7% of the engineering estimates;[9] and whole house retrofit programs achieve just 18% of the engineering estimates, while other programs such as those for dishwashers and clothes washers generate increases in energy consumption. However, programs upgrading pool pumps, water heating, and audits generate larger savings than the ex-ante estimations. This indicates that the bias of engineering estimates can go both ways, underscoring the importance of conducting EE evaluations with actual meter or billing level data.

In terms of program heterogeneity, we find that the level of the financial incentives matters. For example, we find larger savings among products with cost-sharing (e.g. lighting rebate programs) than products given away for free (e.g. free LED light bulbs). Cost-sharing (< 100% subsidy) is more effective than free distribution in the case of lighting programs as the impact largely depends on recipients' usage behaviors. Indeed, it is possible that free light bulbs

8 The covariates for matching are location (X Y coordinates), single vs. multi-family housing, vintage, square footage, etc., and socio-demographic variables from the census block group data (i.e. income, population density, education, etc.

9 The estimated coefficients for lighting programs are not significant in our most conservative specification.

are not installed by households, in contrast with light bulbs purchased through rebate programs. Yet, we do not find differences between programs giving incentives to the distributors and those giving incentives directly to the end-users. The result highlights the possibility of improving program effectiveness by choosing the appropriate delivery mechanism for each products or technology.

In terms of cost effectiveness, we find that pool pumps, and water heating are the most cost-effective type of EE upgrade (cost $0.015 per kWh saved, and $0.001 per kWh saved, respectively). HVAC and whole house retrofits are relatively costly measures to reduce electricity consumption since they only provide savings during certain times of the year ($0.19 per kWh saved, $1 per kWh saved, respectively). Refrigeration programs are also promising given they generate large overall savings and are cost-effective. The direct program cost for refrigeration upgrade is $0.03 per kWh saved—quite comparable to a demand response behavioral program ($0.025 per kWh saved) (Allcott and Mullainathan 2010). However, our findings also suggest that policymakers proceed with caution when designing subsidy programs to upgrade appliance (dishwashers and clothes washers), building shell, HVAC, and whole house retrofits. The first two could lead to perverse upgrades that increase electricity usage, and the last two only provide savings at certain times of the year.

The paper proceeds as follows. Section II locates our study in the current literature. Section III provides the background for the study of energy efficiency programs. Section IV presents our data and empirical strategy. Section V presents our main results on overall program impact. Section VI details heterogeneous estimates. Section VII discusses program cost effectiveness. Section VIII presents robustness checks. Section IX provides a concluding discussion.

II. Literature Review

There is growing empirical literature on the effectiveness of EE programs. However, due to privacy law applicable to accessing individual utility data, large-scale evaluations using credible quasi-experimental design are still rare. Estimates of savings in EE program impact evaluations often rely on model simulations and extrapolations, and rarely incorporate credible pre and post energy consumption data (Qiu and Patwadhan 2018). In addition, most government reports lack credible statistical approaches to overcome selection issues. Consequently, the results from these evaluations are not easily generalizable.

In the recent years, researchers have started to use smart meter and billing data to evaluate realized savings. In Table 1, we provide a list of the most recent academic research evaluating residential energy efficiency programs using micro data.[10] The list indicates that most studies show lower realized savings than the ex-ante engineering predictions, ranging from 25% to 79% of the ex-ante projections. This indicates the need for adequate evaluation methods to better understand energy savings and program effectiveness.

***

[Insert Table 1 Here]

***

Despite the growing literature evaluating residential energy efficiency programs, few studies use large-scale household-level data with randomized controlled trials or quasi-experimental designs. Further, most of these evaluate only a single program without controlling for potential participation in multiple programs, leading to possible biased estimates (See Table 1). For example, Fowlie et al. (2015) study the effectiveness of the Weatherization Assistance Program (WAP) among 34,161 eligible low-income households in Michigan in 2011-2014. They find that WAP reduced monthly energy consumption by 10-20%, but achieved only 40% of the projected savings. Based on two Wisconsin residential EE programs, Allcott and Greenstone (2017) find that actual savings amounted to only 58 percent of projected savings. Those programs did not yield positive social welfare in terms of cost-effectiveness. Boomhower and Davis (2017) focus on a program providing air-conditioning rebates in California. They find that the cost savings from AC units are underestimated if we ignore the timing of the savings. Because these studies focus on individual programs, and because these programs are in different locations, it is difficult to compare the magnitude of the savings across programs.

Davis et al. (2014) is the only study that uses large-scale data comparable to our study. They analyze 1.9 million Mexican households that replaced their refrigerators within 2 years.

10 We cross-checked our list and information referenced from the most recent review paper done by Chiu and Patwardhan (2018).

They found that the program has a positive impact on reducing electricity usage (reduced by 8 percent), but that the actual savings accounts for only one-quarter of the predicted savings. However, energy efficiency markets in Mexico are quite different from the U.S. Therefore, the question remains with respect to the effectiveness of similar programs in the U.S. Additionally, we analyze multiple types of appliance upgrades and can provide a richer context for policy-making.

Furthermore, because utility billing data does not include building-level information, very few studies include information on building characteristics. When they do, there are insufficient numbers of observations to examine building heterogeneity (see Table 1). As building quality explains a significant amount of energy use, it is imperative to include building characteristics to help control for selection bias in program participation. It might also help understanding heterogeneous program impacts.[11] For example, the realized savings of an AC unit upgrade can be greater in a new and well-insulated building, than in an older building not well insulated. It is therefore difficult to extrapolate some of the current literature results without taking into account building characteristics.

Our study contributes the existing literature in three ways. First, we use a large sample of households (more than 11 million unique accounts) with monthly electricity usage in 2010-2014 to construct a rigorous quasi-experimental method to address potential endogenous program participation. Second, we examine all available end-user programs, thereby taking into account overlapping program effects. Third, we analyze heterogeneous program impacts by program types and building types. Although program take-up is often low, our large-scale sample makes it possible to examine heterogeneous program impacts with enough statistical power.

III. Program Background

Our study focuses on all the customer incentive based EE programs administered from 2010-2015 by Southern California Edison (SCE)—one of the largest utilities in the United States, providing electricity to more than 14 million people. During that period, SCE provided

[11] Building quality have great potential to improve energy savings. For example, Asensio and Delmas (2017) examined the impact of certified green buildings on energy savings, and they found that certified green buildings (e.g. LEED building) can save up to 30% of energy consumption, compared to non-certified buildings.

financial incentives to customers for upgrading their home with more energy efficient products, such as lighting, pool pumps, refrigerators, and other appliances. To obtain the rebate, SCE customers typically need to upgrade to a more energy efficient product in their homes during the program implementation period, and then apply for the rebates by mail or online. For example, single-family home owners can apply for rebates through the Home Energy Efficiency Rebate (HEER) Program—the largest residential program based on expenditure (CPUC 2015), and multi-family owners can apply through Multifamily Energy Efficiency Rebate Program for various upgraded appliances, such as HVAC, lighting, pool pump, and fridge. Other programs provide larger financial incentives for EE upgrades. For example, the Comprehensive Mobile Home program focuses on promoting EE technologies among mobile home owners and provides direct install for lighting and HVAC upgrade at no charge.

Table 2 shows the name and the number of SCE households enrolled in different EE upgrade programs. We also present the detailed program implementation strategies in Appendix B. There are 11 types of products based on SCE's categorization, namely appliance upgrade (i.e. dishwasher and clothes washer), consumer electronics, HVAC retrofit, lighting, pool pumps, refrigeration (i.e. fridge and freezer), water heating, audits, whole house retrofit, and building shell. Note that households may claim multiple incentives to upgrade their homes through various programs. To be consistent with SCE's program design rationale, we focus our analysis on product categories. As Table 2 shows, among those products, lighting and refrigeration have attracted larger participation.

***

[Insert Table 2 Here]

***

EE programs might use different levels of financial incentives (as shown in Table 3). They can provide the products for free or subsidize a certain percentage of the cost (i.e. rebate). SCE also provides incentives to different recipients. For example, incentives given to contractors or distributors are called up-stream or mid-stream incentives. Down-stream incentives are given directly to households, the end-users.

***

[Insert Table 3 Here]

***

IV. Data and Empirical Framework

Data Description

We combine four datasets to understand the effectiveness of energy efficiency programs among residential households in Southern California. We focus on programs where households could claim financial support between 2010-2015.[12] To evaluate the impact of the programs on energy usage, we use monthly electricity consumption measured at the household account level.

Electricity usage and program participation data are extracted from the LA Energy Atlas, a relational database that enhances understanding of energy usage across LA County (Pincetl and LA Energy Atlas Development Team, 2015). [13] Key data for EE program participation are from both the California Public Utilities Commission (CPUC) and SCE, the regional electricity utility. Additional building level characteristics information comes from the Los Angeles County Assessor's property dataset, as well as socio-demographic information from the census database. See Table 4 for more details regarding the source and coverage of our data.

***

[Insert Table 4 Here]

***

EE Program Data. Program participation data identifies residential energy efficiency programs implemented in SCE service territories during 2010-2015. Note that 2015 electricity consumption data is unavailable to us, so for the program impacts on electricity savings, we only analyze the information from 2010 to 2014. Yet, the program information from 2015 is useful to

[12] Program related information (ex: installation date, rebate amount, etc.) was recorded. However, we do not analyze other programs that do not directly involve end-users' action: for example, training and education programs to promote EE upgrade, subsidy programs for developing new EE technologies and standards, etc.

[13] www.energyatlas.ucla.edu

prevent us from selecting any future EE participants into our matched control group. More than 80% of our observations have at least 12 months of data for the analysis.

Each time a household claims rebates or direct financial support for an upgraded equipment/appliance, there is a record that documents related program information, such as the installation date, rebate amount, type of product, and predicted savings. Based on the installation date claimed in the form, we generate a program participation variable reflecting the starting month of the upgrade for each household and each product. All subsequent monthly information for an EE participant is considered a "treated period" for that certain program/measure once a household has participated in a program.

The EE program participation data cover 506 cities.[14] We illustrate the participation information in Table A1 and Figure A1 of the Appendix. Comparing EE participation rates across income categories shows more participation by higher income areas (the top income quantile). The take-up rate is 11.6% for the highest income quartile, and 5.5% for the lowest income quartile, respectively. Orange County has higher EE adoption rates (12%) than the other counties. Los Angeles and Imperial County have lower adoption rates (around 6 percent) than other counties. This indicates the need for a credible quasi-experimental design to mitigate this self-selection bias when evaluating program effectiveness.

Electricity Usage Data. Account-level electricity billing data are available from January 2010 through December 2014 across the SCE service territory. Our data contains more than 11 million unique accounts with monthly electricity usage data for 2010-2014. The panel data is unbalanced because households may move in and out of the building, or even the area so they may not have a complete record throughout the study period. Yet, more than 80% of our observations have more than 12 months of data for the analysis.

The unit of analysis is the combination of the household utility account and the building. If a household has moved to a new address, even while carrying over their same utility account, we generate a different ID. This approach accounts for the potential that the same household may

[14] Including Irvine, Lancaster, Santa Ana, Palmdale, Valencia, Aliso Viejo, Orange, Rancho Santa Margarita, Corona, Fullerton, Long Beach, Moreno Valley, Lakewood, Newhall, Costa Mesa, Tustin, Mission Viejo, Los Angeles, Torrance, Saugus, etc. However, the data do not cover information for LA City, which is served by the Los Angeles Department of Water and Power (LADWP), a separate utility.

consume electricity differently due to building characteristics, such as vintage and square footage. Also, households may not bring their appliances to a new building, so they are no longer considered treated once they moved.

We further illustrate the average consumption pattern by calendar month of the year for EE participants and non-participants in Figure 1. It is evident that electricity usage exhibits a seasonal pattern. As shown in the figure, EE participants tend to consume more (ranging from more than 600 kWh per month to 900 kWh per month) than non EE-participants (ranging from around 500 kWh to almost 800 kWh per month). Even though non-participants have reasonably similar consumption pattern to EE participants before matching, this figure shows that there is room for better selecting the comparison group. This difference between EE participants and non-participants motivates us to control for time-invariant characteristics using household-by-month-of sample fixed effects. We also recognize that many variables (such as building vintage, size, etc.) that determine program take-up, may still be correlated with electricity consumption between EE participants and non-participants. As the literature shows that regression models may perform poorly with small covariate overlap (Dehejia and Wahba, 1999, 2002; Glazerman et al., 2003), we use a matching method to improve our selection of the counterfactual group based on all the plausible observables to improve covariate overlap. The empirical strategy section explains our method in detail.

***

[Insert Figure 1 Here]

***

Building Characteristics Data. Building characteristics used in this analysis include use type (i.e. single-family housing, multi-family housing, condominium, etc.), square footage, building vintage, climate zone, whether the house is occupied by its owner as the primary residence, and whether the household is registered under CARE/FERA, energy discount programs for low income households, and whether the household is identified as having a swimming pool or not. The descriptive analysis of the key variables is shown in Table 5. For LA County, building use type and ownership information is sourced from the 2016 county assessor office's parcel database that is publicly available on the LA County GIS portal website. For all other counties, information is from a standardized parcel database created by the Southern

California Association of Governments (SCAG). Through a multi-stage geocoding process developed for the LA Energy Atlas, individual account addresses are associated with parcel boundaries. This process links each utility account, and its associated utility consumption records, with the building attribute information available for the various parcel database sources.

***

[Insert Table 5 Here]

***

Socio-demographic Data. We use US census data to obtain socio-demographic information. The census information is taken from the American Communities Survey (ACS) 2006-2010 estimated at the block group level. The variables of interest include median income, population density, poverty, ethnicity, education, the percentage of homeowners, and occupancy rate.

Empirical Strategy

The program impact can be identified by comparing the change of electricity consumption over time between participants and non-participants. The challenge for the evaluation is that many other factors may impact households' program participation decision as well as their energy usage. In order to rule out factors that may confound the program impact, we first find a set of households that are similar to our program participants but have never participated in any of the programs. We use a matching method to find this relevant comparison group based on observable variables such as location, building characteristics, and demographic variables—all of which are determined to be important regarding decisions on program uptake in the literature. We did not use any outcome variable to construct our control group so as to retain the objectivity of our design, as recommended by Rubin (2007) when drawing casual inference from observational studies. This pre-match method can improve the covariate overlap to reduce bias before running the panel regression models with various household fixed effects (Dehejia and

Wahba, 1999, 2002, Ho et al. 2007, Stuart 2010).[15] This covariate overlap improvement using matching before panel fixed effects regressions is also evident in the empirical setting (Alix-Garcia et al. 2015).

We pre-match households that are program participants in any year between 2000 and 2015 with households that have never participated in any EE programs using Mahalanobis Distance Matching with replacement. This matching scheme uses the data from over 10 million non-participants and finds the nearest neighbors to our program participants based on several covariates that were not previous available in the literature. The idea is to establish an appropriate counterfactual sample of households that are as similar to our participants as possible, but have never participated in any of the energy efficiency programs in our study. To conduct the matching, we require exact match within the residential use type (e.g. single family housing, multi-family housing, condominium, combined residential), vintage bins (built before 1950, between 1950 and 1978, between 1979 and 1990, and after 1990), square footage bins,[16] climate zone, and fuzzy nearest distance matching over other covariates include owner-/renter-occupied,[17] whether they participated in any low-income electricity discount program, either California Alternate Rates for Energy (CARE) Program, or Family Electric Rate Assistance (FERA Program), and whether they had a pool.[18] We also include x y geographic coordinates to make sure that we can minimize this extra dimension to improve our location matching.[19] Our contribution is to use both socio-demographic and building-level information in selecting our counterfactual group as the existing literature uses covariates from socio-demographic

[15] Nonetheless, we are aware that the matched comparison group is not drawn from a randomized experimental setting. There may still exist differential time trends in electricity consumption between participants and non-participants. We address this concern in the robustness check section.

[16] Because square footage has a very wide range, we bin them based on 100 percentiles.

[17] Based on property tax records, we can identify those who are the building owner and at the same time live in that building as their primary residence.

[18] Pool ownership information is limited and incomplete. Therefore, for those who do not have pool information, we simply impute the missing values that simulate the distribution. We can treat this process as if we use this extra pool information to improve the matching method without throwing away observations. We have also tried matching without this pool ownership information, and the result is consistent.

[19] Since we use x-y coordinates in matching, some may worry about spillover effect—non-participants who live adjacent to our treated participants may become more energy saving; for example, they learn more about energy-related knowledge. Nonetheless, this potential spillover effect will make our program impact under-estimated which means that our identified program impact, if any, is even stronger in the absence of spillover. Also, we use similar matching covariates but without x-y coordinates, and the results are consistent, indicating small spillover effect.

information (mostly at the census blockgroup level) or past electricity usage to find the counterfactual. Very few studies have building-level information (see Table 1), which is considered as more important determinants in explaining energy efficiency retrofit investment than socio-demographic characteristics (Trotta 2018).

After matching, we use panel regression models to identify impact at the household level. By controlling for household fixed effects or household-month fixed effects, we eliminate unobservable household-level characteristics that do not change over time. Household size, political affiliation, and environmental attitude, are all examples of time-invariant variables that may affect program participation as well as electricity consumption. Controlling for household fixed effects in the model can rule out those characteristic differences between participants and non-participants. We also control for time fixed effects for any unobservable factors that may confound the program impact.

In the post-matching estimations, we rely on the difference-in-difference (DID) technique to identify program impact. First, we determine the difference in electricity use between program participants and non-participants to account for the systematic electricity usage difference (first "difference"). Then we calculate another difference to compare electricity usage before and after retrofits for participants as well as non-participants (second "difference").

To evaluate the overall treatment effects for residential households, we estimate panel regressions (equation 1) with various fixed effects on the pre-matched sample with the following specification:

$$\ln(Energy_{imt}) = \beta EE_{imt} + \alpha_{im} + \gamma_{mt} + \varepsilon_{imt} \tag{1}$$

where EE_{imt} is an indicator variable for identifying household *i* switches from zero to one when that household joined any EE program in month *m* year *t*. We use the month when households installed the product to determine their status as a participant of the program. To understand energy usage, we use $\ln(Energy_{imt})$, which is the natural log of energy usage (in kilowatt-hours (kWh)) for household *i* in month *m* year *t*. Because of the large variation of this variable (for example, extremely heavy energy users in Beverly Hills), we use a natural logarithm to smooth the consumption variable at the extremely higher end of the distribution. This method also eliminates noisy observations that have zero electricity consumption, as even keeping a refrigerator on will take at least 20 kWh a month. Unfortunately, we do not include natural gas consumption data as they are unavailable to us.

Household-month fixed effects (α_{im}) take into account unobservable household characteristics in a certain month that may affect energy usage. That said, we create 12 separate fixed effects for each household. These household-month fixed effects address unobservable time-invariant differences in attributes between program participants and non-participants. In this context, the time-invariant unobservables refer to those characteristics that may affect program participation, and at the same time affect electricity usage; for example, a person's environmental attitudes, household size, and political views. Our model controls for these, as long as they are time-invariant. Household fixed effects within the month in a year also take away the systematic seasonal pattern of household's electricity consumption. Identification comes from the within household-month variation. For example, if a household joins a program on June 2010, we differentiate all the months before the participation and the corresponding months after the participation. This means that we are comparing January 2010 (before) to January 2011 and January 2012 (after), February 2010 (before) to February 2011 and February 2012 (after), March 2010 (before) to March 2011 and March 2012 (after), and so on.

We also include month-year time fixed effects (γ_{mt}) to control for economic or administrative shocks in each time period. During certain times of the year, some places may experience economic shocks and need to diverge more sources to other welfare programs instead of environmental programs; these unexpected economic shocks may also correlate with energy consumption in the region. As program impacts may be overstated with administrative capacities varied by time, γ_{mt} difference out those confounding administrative and economic shocks.

To better understand what type of program works, we further analyze program effectiveness by dividing programs into different types of products. We estimate equation (2):

$$\ln(Energy_{imt}) = \beta_1 EEProduct1_{imt} + \beta_2 EEProduct2_{imt} \ldots \alpha_{im} + \gamma_{mt} + \varepsilon_{imt} \quad (2)$$

where $EEProduct1_{imt}$, $EProduct2_{imt}$ ... are indicator variables for identifying household *i* switches from zero to one when that household join an EE program upgrading their home appliance/equipment (i.e. consumer electronics, HVAC retrofits, lighting, etc) in month *m* year *t*. Our main goal is to test the mean change in electricity consumption associated with the EE product upgrade (i.e. parameters $\beta_1, \beta_2 \ldots \beta_{11}$). We are interested in testing the effectiveness of the program for different end-use products (whether $\beta_1 < 0$; $\beta_2 < 0 \ldots \beta_{11} < 0$). There are mainly eleven products delivered with financial incentives including appliances, consumer electronics, HVAC, lighting, pool pump, refrigeration, water heating, audits, whole house retrofits, building shell, and other equipment.

V. Main Results

Summary Statistics

Table 6 shows summary statistics of various attributes among EE participants and non-participants in our total sample and matched sample. We examined differences between the characteristics of EE participants and non-participants. Participants are more likely to own their homes, tend to live in newer buildings, in areas with lower population densities and higher incomes with lower poverty rates. Participation rates are higher among white and Asian populations and lower among African American and Hispanic populations. Participants also tend to be from neighborhoods with more highly educated populations. The findings all indicate that participation correlates with greater access to resources, and further justify our matching method to improve covariates overlap.

***

[Insert Table 6 Here]

***

To test covariate balance, we calculate the normalized differences in means between EE participants and non-participants in the whole sample, and that between EE participants and matched non-participants (See Table 6). These figures are calculated by the difference in means between the treatment and control groups divided by the square root of the sum of variances for both treated and control groups. The idea is to calculate the standardized differences in various attributes, so we can compare whether our matching method improves the similarity of the control and treatment group, as this is the most common way to diagnose covariate balance (Rosenbaum and Rubin 1985; Stuart 2010).

We further illustrate this comparison in Figure 2. As Figure 2 shows, our matching improves covariate balance in almost every aspect. Most covariates have a smaller normalized difference in means with the exception of the Asian population and the occupancy rate. With matching, not only has the covariate balance greatly improved, but also all the normalized differences in means are also smaller than 0.25 standard deviations, which is the suggested rule of thumb in the literature (Rubin 2001; Imbens and Wooldridge 2009).

***

[Insert Figure 2 Here]

***

Overall Treatment Effects

We present the overall treatment effects based on equation (1) in Table 7. The results indicate that the residential energy efficiency incentive programs reduce overall electricity usage by 4 percent, or 25 kWh per month. This is equivalent to 311 kilowatt hours of annual savings. The adoption rate for any energy efficiency upgrade is about 8% for the five-year duration. So if we extrapolate this number into 11 million households, this indicates that the overall program impact is equivalent to 54 gigawatt hours in annual savings (311*11,000,000*8%/5= 54,736,000 kWh).

We present the graphical analysis of the overall treatment effect through another perspective (See Figure 3). We estimate an alternative version of equation (1) but instead interact the EE participation variable with indicator variables representing quarters before and after the take-up time. We plot the coefficients and 95 percent confidence intervals for each quarter relative to the time of EE take-up based on the following event study regression.

$$\ln(Energy_{imt}) = \sum_{q=-6}^{6} \beta_q \, 1[quarter\ to\ upgrade = q]_{\,imt} + \alpha_i + \gamma_{mt} + \nu_{imt} \qquad (3)$$

Where $1[quarter\ to\ upgrade = q]_{\,imt}$ is the indicator function representing the relative time to program participation. For example, q=-6 indicates 6 quarters prior to the program participation; q=6 for 6 quarters after EE upgrade. The excluded category is q=0, indicating the exact month of EE take-up. This graph result indicates program impact relative to the month of EE participation. The coefficients are estimated using the matched sample.

Figure 3 displays overall electricity reduction after program participation, and the overall savings seem to persist over time after the second quarter. There is no treatment effect prior to the program participation. However, a slight time trend between participant and non-participants prior to the program participation is exhibited. We address this concern in the robustness check section. The results are similar when controlling for this upward time trend.

***

[Insert Figure 3 Here]

***

Individual Product Effects

Beyond the overall positive program impacts, it is crucial to examine which programs deliver the most savings. Table 7 presents the main coefficient estimates for various EE upgrade programs based on equation (2). All of the specifications are based on the matched sample for estimating the average treatment effect on the treated. Columns (4) is based on specifications with the most conservative fixed effects—household-month fixed effects and time fixed effects as in our estimation equation (2). Columns (1) and (2) include alternative fixed effects—column (1) is with household and county-month-year fixed effects; column (2) is with household and city-month-year fixed effects. Columns (1) to (3) are clustered at the household level to control for serial correlation. In column 4, we also cluster the standard errors at the building level to control for spatial and serial correlation. As standard errors clustered at the building level yield more conservative results, we use this specification as in column (4) in the rest of the regressions.

***

[Insert Table 7 Here]

***

Overall, the magnitude and the significance of the results do not change much, especially for those products that deliver the most savings. For example, pool pump programs, on average, deliver 11 to almost 13 percent of savings; programs that give incentives on refrigeration generate 6-7 percent of savings, on average. However, in the case of lighting programs, we find that the magnitude of savings is smaller and insignificant in the more conservative specification (columns 4) than in the less conservative specification (columns 1 and 3).

The whisker plot (Figure 4) highlights the results based on our preferable specification in Table 7 Column 4—including household-month and time fixed effects and our most conservative standard error estimation. A negative number in the figure means that program participants decreased electricity usage. Based on the results from all the multiple statistical models, pool pump programs yield the highest savings. Households participating in these programs, on average, reduce their energy consumption by 11-12 percent. The result accounts for seasonal patterns: pool pumps may be highly utilized in the summer, making those programs more attractive during certain months of the year. In this case, we may under-estimate the program impact (with small savings) because electricity usage may suddenly go up right after upgrading the pump. As we control for household-month fixed effects, we rule out this selection in month factor by comparing electricity savings in the same month of the year, before and after program participation. Other effective programs for reducing electricity consumption include incentive for upgrading refrigeration (including refrigerator and freezer). Households who have new efficient refrigerators or freezers reduced their electricity consumption by 6 percent, on average.

***

[Insert Figure 4 Here]

***

Lighting programs result in relatively small savings—0.3 to 0.7 percent reductions, and statistically indifferent from zero in our most conservative estimate.[20] The result could be attributable to any of the following three factors. First, lighting may constitute only a small part of household electricity consumption. Second, some light bulbs may not be installed when they are given away for free. Third, light bulb upgrades may have been adopted anyway without those programs. We will not be able to directly verify the third mechanism, but we investigate the second mechanism indirectly in the next section by examining program heterogeneity.

Other programs do not yield large impacts in terms of electricity savings, or result in increased consumption. These include programs that promote appliance upgrades (mostly dishwashers and clothes washers), programs that incentivize HVAC retrofit, whole house retrofit, and building shell upgrades. This could indicate rebound effects, which happen when households increase their electricity consumption with more efficient appliances. It is also possible that some households, when upgrading, may have chosen larger appliances, which may lead to increased energy consumption.[21] There may also be cases where households, when made aware of being able to save energy, spend less effort on energy conservation (Asensio and Delmas 2016).[22] For example, after upgrading to a more energy efficient product, people do not unplug their charging devices, nor turn off unused lights. We do not have direct evidence in the data to determine the most likely explanation. However, we investigate this issue in the next section through an analysis of heterogeneous effects.

Water heating, audits, and consumer electronics programs yield less conclusive results across models. Results from the water heating and audits programs yield about one to two percent savings.[23] Consumer electronics programs yield insignificant results. The lack of savings

[20] This zero saving result is robust after dropping light bulb incentives given away at the distributors/retailers because after interviewing the program managers at SCE, they worry that upstream/midstream light bulb incentives may not be correctly recorded in the program data.

[21] Houde and Aldy (2017) find that rebate programs induce a potential income effect that EE participants upgrade to a more energy efficient, yet larger appliance.

22 Asensio and Delmas (2016) observe residential households' dynamic energy behaviors at the appliance level through high-frequency smart-meter technology, and they find that when households save energy by turning lights off, there is potential associated rebound effect of increasing energy usage by plug load and heating and cooling.

[23] Water heating upgrade is mainly through providing free water saving kits (ex: low flow showerhead, faucet aerator, etc.) through a so-call Home Energy Efficiency Survey Program.

resulting from consumer electronics EE programs may be due to the small sample size – only 652 customers have ever participated in those programs.

We run the same regressions as in equation (2) by month to see how the results may differ in different seasons. We present our results in Figure 5 and Appendix Table A2. Consistent with our results in Table 7, the pool pump upgrade and refrigeration upgrades lead to constant savings in different seasons. HVAC programs have a stronger seasonal effect—positive savings in the summer while no or even negative savings in other months. This result is consistent with the literature. HVAC programs deliver savings during different months of the year or even different times of day. The result highlights the concern that the impact of HVAC programs may be limited by the behavioral responses of the users, while those more "passive" programs, like pool pump and refrigeration programs, in general lead to more persistent savings throughout the year. Whole house retrofit programs exhibit seasonal saving patterns similar to HVAC programs, but with larger confidence intervals and smaller savings in the summer months. However, the result is less conclusive for the whole house retrofits programs as they start after 2013.

Comparison with Engineering Predictions

In order to quantify our estimates and compare them to the ex-ante engineering estimates recorded by SCE, we estimate equation (2) using monthly electricity consumption in kilowatt-hours (kWh) as the outcome variable. The ex-ante saving predictions are based on SCE's original recorded saving projection. Those ex-ante predictions are derived from The Database for Energy Efficient Resources (DEER), a publicly available tool for predicting energy savings for various energy efficient technologies and measures. For each specific upgrading product, we calculate its average predicted monthly savings based on SCE's assigned life-cycle savings for that installed product. The comparison is summarized in Table 8.

***

[Insert Table 8 Here]

***

As Table 8 shows, program impacts using actual billing data are in general inconsistent with SCE's ex-ante predictions. We find that this overstating projection is particularly strong for lighting, where the estimated savings are 13 times more than the actual savings. Some programs even show potential rebound effects—incentive programs for retrofitting building shell and other appliances (e.g., dishwasher, clothes washer) lead to more electricity consumption. The measures that are on target or even out-perform the ex-ante saving predictions are pool pumps, water heating, and audits.

VI. Heterogeneous Effects

Program Financial Incentives

We further investigate the program impacts by the level and delivery mechanism of the financial incentives.

Table 9 reports estimates from four separate regressions based on the program impacts interacted with the way subsidies were distributed, and the level of subsidy. For example, some programs offer indirect financial incentives to upstream actors (i.e. product manufacturers) and midstream actors (i.e. retailers or service providers) to promote energy efficient upgrades, while some programs offer so-called downstream incentives that target end-users through mail-in or in-store rebates or discount.

***

[Insert Table 9 Here]

***

As Table 9 shows, free give-away lighting programs (results in row 6, 9) generate smaller savings, compared to those that provide partial financial incentives.[24] On average, lighting rebate programs lead to 3 percent significant savings, while free lighting programs lead to only 0.3-0.4 percent savings. The result stays the same when controlling for the quality of light bulbs using the ex-ante predicted savings (Column 2 in Table 9). This indicates larger impacts among products with cost-sharing delivery methods as compared to the same type of products given away for free. This is in line with the public finance literature that has identified several positive

[24] The exception is HVAC programs. Results for up/midstream HVAC and HVAC rebate programs are not significant because there are very much fewer observations in up/midstream HVAC programs. We cannot conclude whether this statement applies to HVAC.

effects of cost sharing: first, a selection effect which cost-sharing helps select those who need the product more and therefore use more;[25] second, a psychological effect that people exhibit behavioral bias by using the product more if they pay for it (similar to sunk cost effect) (Thaler, 1980; Arkes and Blumer, 1985); third, a signal effect that people view the product as having higher quality, thus encouraging their usage (Bagwell and Riordan 1991; Riley 2001).

Therefore, cost sharing, while potentially reducing the demand for energy efficiency investment, may induce a positive selection of households who need or value the product more, and thus use the product more appropriately. The result highlights the possibility of improving the effectiveness of the program by choosing the appropriate delivery mechanism for different products. The result is not very conclusive comparing the up/midstream programs vs. the end-user rebate programs.[26]

Household and Building Characteristics

We also conducted sub-group analysis. We estimate equation (2) by income quartile, vintage, square footage, and climate zone to understand where programs deliver the largest/least savings. This analysis can help policymakers better target programs and areas that deliver the largest impact.[27]

We find that the magnitude of the coefficient for pool pump and refrigeration programs are consistent with the main result in Table 7 Column 4 based on the sub-group analysis. Therefore, we focus our analysis of heterogeneous effects among programs that have potential rebound effects—EE participants seem to use more electricity after upgrading, compared to non-

[25] In contrast, free lighting programs may give free light bulbs to those who do not need them; for example, people who just replaced their light bulbs, or those who would never throw out perfectly good light bulbs even though they save energy.

[26] Pool pump programs incentivizing end users seem to generate slightly larger savings than incentivizing up/mid-stream manufacturers and contractors. However, the difference may not be considered large as to economic meanings.

[27] The difference between engineering estimates and actual savings can also arise because some EE subsides are "non-additional," meaning that participants would have bought the EE upgrade anyway without the program. Boomhower and Davis (2014) use regression discontinuity to calibrate how large this effect could be and found that half of the EE participants would have done the upgrade with no subsidy. We could capture this concern in our control group. For example, people who have never participated in an EE program may actually have upgraded their lights, leading to a lower "treatment effect on the treated," using the DID method. Therefore, our model identifies the "additional" program impact, while engineering prediction identifies the savings purely from the equipment alone.

participants. Then we try to compare electricity usage among different income quartiles, square footage quartiles, and vintage sub-groups.

Table 10 compares program impact between those who live in a lower income neighborhood (below median income) and those who live in a higher income neighborhood (above median income) using median income information from the census block group data. We do not see significant differences by income groups across various products, except for audits.[28] Audits programs lead to slightly lower savings for households in lower income neighborhood (2% savings) than households in higher income neighborhood (3% savings) although this difference is economically miniscule. However, since income is identified under a coarse block group information, rather than account or building level, we need to take this result with a grain of salt.

***

[Insert Table 10 Here]

***

Table 11 compares program impact between those who live in a larger home (1st and 2nd square footage quartile) and those who live in a smaller home (3rd and 4th square footage quartile).[29] Most of the comparisons by the size of the building do not yield economically significant differences. The only exception is the audits programs. Audits lead to around 5 percent savings for large buildings (3rd and 4th quartile), and 1 percent savings for smaller homes (1st and 2nd quartile). This result may simply be due to the larger savings potential in retrofitting a large building.

***

[Insert Table 11 Here]

***

[28] Even though the coefficients of HVAC, lighting, water heating, and whole house retrofit look slightly different by income group, the differences are not statistically different. We also find a consistent pattern by income quartile.

[29] For multi-family housing and condominiums, we cannot clearly identify each account's exact square footage---we can only identify the building structure they live in based on geocoding their account address to match with assessors' tax database. Therefore, we use single-family housing for this sub-group analysis.

The other part of Table 11 compares program impact between those who live in a home built before 1978 and those in a home built after 1978. We chose 1978 as the cut-off year because California Title 24 Building Energy Efficiency Standards were established at that year.[30] The most interesting results here concern HVAC and whole house retrofit programs. For HVAC, coefficients are positive before 1978 and negative after 1978. This indicates that HVAC program participants in older buildings use more electricity after joining the program, compared to their non-participant counterpart. It could be that energy efficient HVAC and building efficiency may complement each other. For example, Liang et al. (2016) found that some initial building attributes may affect the impact of retrofits. For example, HVAC duct sealing retrofits, a type of popular retrofitting in our data, can be more effective with better roof insulation. We also find suggestive evidence in our data that those who have participated in both HVAC retrofit and whole house retrofit programs reduced their electricity consumption by 8 percent more than those who have simply done a HVAC retrofit.[31]

Regarding whole house retrofitting, the coefficient is negative before 1978 and positive after 1978. This result indicates that older homes benefit from the retrofit program. This is consistent with engineering understanding that EE investments in older buildings may yield larger saving potentials. However, newer homes increase their consumption after the retrofit. This surprising result might be explained by an increase in the size of the appliances installed during the retrofit.[32]

The results for the whole building retrofit programs are the opposite. Participants in older buildings after retrofitting consume approximately 2-3 percent less electricity, while participants in newer buildings use more electricity after participating in the program. This heterogeneous effect in whole building retrofits indicates that older buildings may have larger saving potential. Nevertheless, we recognize the fact that all whole house building retrofits have potential to improve natural gas savings. However, we do not have access to natural gas data for further

[30] There may be a lagged effect for Title 24 implementation that attenuates this comparison. However, this attenuation effect will make our results even stronger if we find drastic heterogeneous program effects comparing buildings constructed before 1978 with those built after 1978.

[31] We only have 46 households participated in both HVAC and the whole house retrofit programs, so we are not able to estimate further by the building type.

[32] Based on anecdotal evidence from one anonymous SCE program manager, they have also seen cases where people increase their size of homes when doing a whole house retrofit.

analysis. This limitation may cause us to underestimate the overall energy savings from the whole house retrofit programs and the magnitude of this underestimation will depend on the use of natural gas services in homes.

Table 12 examines the impact of energy efficiency programs by climate zone. A climate zone is defined based on its weather pattern. The overall results are consistent with the main results. The saving estimates of HVAC and whole house retrofits are heterogeneous as we expect them to be more sensitive to the local weather pattern. This result is in line with Figure 5, which shows the seasonal impact of HVAC and whole house retrofits. For example, HVAC retrofits deliver 1% savings in climate zone 10, which requires higher demand for energy needed to heat a building (1678 heating degree days for the representative city in the zone), while they deliver a 5% increase in electricity consumption in zone 6 where heating is not in high demand (742 heating degree days for the representative city in the zone). The saving results from the whole house retrofit programs also differ from zone to zone.

VII. Cost effectiveness

We evaluate the program cost-effectiveness by product in Table 13. Column 1 shows the mean annual savings by product based on the result in Table. From the program data, we calculate the average direct program cost for each upgrade. Direct program costs are the total payments associated with the upgraded EE product, mostly subsidies paid to the participating households, [33] and do not include indirect program costs such as administration, advertisement and training. We calculate the cost per kWh saved using the program direct cost for each product and then divide it with the total life-cycle savings. For the total electricity saved, we use annual savings times the estimated average life cycle reported in the program data at the zero discount rate (Table 9 Column 4). We do not report savings for products that do not yield significant savings.

***

[33] Some retrofits are implemented through a so-called "direct install" program. Direct install programs give subsidies to approved contractors to "directly install" energy efficient products for EE participants. In the program data, the direct program cost may be recorded as labor cost and material cost in this case. Incentives could be direct payment to the utility customers, or through indirect subsidy to the distributor and contractors.

[Insert Table 13 Here]

***

Based on this assumption, we find that pool pump and water heating are the most cost-effective type of EE upgrade programs ($0.015 per kWh saved, and to $0.001 per kWh saved, respectively), while HVAC and the whole house retrofits are relatively more costly ($0.18 per kWh saved, $1 per kWh saved, respectively). The magnitude of those estimates is consistent with the literature. For example, Liang et al. (2017) estimate that cost per kWh saved is $0.434 for residential buildings under a 5% discount rate. This number is comparable to our cost per kWh saved estimate for HVAC with similar assumption and similar type of upgrade ($0.42 in Table 13 Column 7). Estimates from information behavioral programs, such as those from Opower, provide an average cost $0.025 per kWh saved (Allcott, 2011). These figures indicate that while pool pump, refrigeration, and water heating are comparable to these behavioral programs and may yield a more persistent result in the long run; other types of energy efficiency upgrades are less cost-effective than the non-price based behavioral interventions.

As cost-effectiveness analysis is sensitive to the assumptions we make, we also calculate the cost per kWh saved based on different assumptions—with 5% discount rate, 5-year life cycle, and 20-year acceleration rate (Table 9 Column (4) to (7) respectively). This simple calculation allows us to compare cost-effectiveness across programs. For example, under the assumptions of a 5% discount rate and 5-year acceleration, we calculate the cost per kWh saved based on the average program costs per upgrade divided by the net present value of 5-year kilowatt-hour savings. With the most pessimistic estimates, pool pump, water heating and audits cost less than $0.03 per kWh saved. With the most generous estimates, given a 20-year life-cycle savings and zero discount rate, HVAC upgrade costs $0.09 per kWh saved, and the whole house retrofit programs cost $0.697 per kWh saved. We nonetheless recognize the limitation that we quantify the cost-effectiveness based on the stated policy goal, reducing carbon emissions through reduced energy consumption. There may be other cobenefits associated with those upgrades, such as home comfort or health improvement, that are not quantified in the cost-effectiveness analysis.[34]

VIII. Robustness Checks

One important assumption for this difference in difference estimation is that in the absence of the program, the matched non-participants should have a similar electricity consumption pattern to EE participants after controlling for those household-month and time fixed effects—there is a parallel time trend between participants and non-participants had there been no programs. This assumption is not directly testable. Yet, we can assess the robustness of our results by adding a time trend variable—a similar test as in Davis et al. (2014). For EE participants, we assign them the number of months since January 2010, and for non-participating households, the value is equal to zero for all months. This helps us to control for the differential trend between participants and non-participants, if any. We report our results including time trend variables in Appendix Table A3, in linear form, in quadratic form, and in cubic form. The results, especially for those more effective programs (pool pumps and refrigeration) are not sensitive to adding the time trend variable, leading to 13% savings, and 6% savings, respectively, which are the same as our main result.

Table A4 column 2 shows another robustness check by dropping the exact month of participation. This robustness check addresses two potential concerns. First, the actual installation time may take few hours to a month to be effective, so this alternative specification allows a lag for the treatment–similar to the replacement time assumption in Davis et al. (2014). Second, there may be a selection bias that people pick the “peaky” month of the year where electricity usage is the highest to upgrade their house. Even though we think that this bias is unlikely given our household-month fixed effects, there may be an extreme case where, for example, households systematically upgrade their appliances during the record high peak usage month throughout the study period, leading to overestimated saving impact. The results in Table

[34] For example, HVAC and whole house retrofit programs may generate side benefits such as home comfort. However, we did not measure that because these side benefits are not the main policy goals of those programs, and measuring those side-benefits requires additional innovative data and approach. There may also be health benefit associating with less electricity generation and indoor/outdoor air quality and temperature. This type of analysis is out of our scope of analysis as it requires extra modeling and data collection.

A4 Column 2 are consistent with the main specification, confirming that the above issues are not a concern.

We also consider whether those who have solar panels may exhibit unusual electricity consumption patterns in their billing data. We thus validate the main results by dropping those who may have solar panels (results shown in Table A4 Column 3 in the Appendix). In general, SCE offers a net energy metering policy that tracks the net difference between the amount of electricity produced and the amount of electricity consumed. The amount will be directly reflected on the bill. For example, if in a given month, the household produced more electricity than it consumed, there will appear to be a negative number (e.g. -100 kWh) on their bill. Given that households that participate in the EE programs may in the meantime be more likely to install solar panels, we re-ran the same model as in Table 7 Column 4 by dropping those who may have installed solar panels. The limitation is that there are no good data to directly identify households with solar panels at the utility account level.[35] Therefore, we identify households who have any negative number in any of the months during the study period, as those with solar panels. We spot-checked this algorithm by verifying those identified households' addresses and their Google image to make sure that this algorithm is not simply identifying data recording errors. Based on the result in Column (3) Table A4, all the coefficient magnitude and the significance level for all products are consistent with our main specification.

Table A4 Column 4 presents the same estimated regression as the main result in Table 6 Column 4 using only participating households. The idea here is to compare those who have participated earlier with those who have participated later as one may think that those earlier participants are a better counterfactual group than those who have never participated. The coefficients are similar to the main specification.[36] This result indicates that the program impact is mainly determined by the within-household variation. That said, even though we are not able to select the perfect counterfactual through a randomized control trial, our result is not sensitive to this limitation. We ran 11 separate regressions to estimate program impact for each type of

[35] The best available solar panel information is aggregated at the zip code level. Since we already matched our households by the building characteristics at the account level and various social-demographic data at the census block group level, this information at the zip code level will be too coarse to provide useful variation.

[36] Lighting programs are more effective in this specification, but the magnitude is still pretty small as the previous result suggested.

upgrade one by one without controlling for overlapping upgrade. Comparing with our result in Table 7 Column 4 (presented here in Column 1 Table A5), we find that most results are consistent, except for lighting and water heating programs. The program impacts are slightly overestimated if we do not control for other product types. This result again stresses the need for controlling for overlapping program impact—one of our important contributions in the paper.

IX. Conclusion

Given the large amount of public funding spent on promoting energy efficiency upgrades, our analysis adds credible empirical evidence about the effectiveness of ongoing EE programs. Overall, our results, based on all programs providing cash incentives in SCE service territory, point to around 4 percent savings. However, some programs perform better than others. For example, programs subsidizing pool pumps and refrigeration lead to larger savings (12% and 6%, respectively), and this positive program impact is persistent throughout the year. Other programs giving incentives for appliance upgrades (mostly dishwasher and clothes washer), whole house retrofit, and building shell have concerns regarding rebound effects.

In addition, the results highlight the importance of incorporating measured electricity consumption in program evaluations, as we found discrepancies between engineering estimates and actual measured savings. For example, the CPUC's overall impact evaluation report claims that lighting programs could deliver large savings based on engineering estimates. However, this is not what we find using our data on electricity usage. Some of the overestimations in engineering models may be explained by the way the programs were implemented. For example, programs giving away light bulbs are associated with very noisy and non-significant saving outcomes.

It is also notable that some programs cause greater usage of electricity. Cash incentives for HVAC retrofit and dishwasher or clothes washer upgrades may induce a behavioral rebound effect where consumers use more energy or spend more on larger or more powerful units when their per unit cost of energy service is cheaper. Unfortunately, we do not have pre-program appliance ownership and usage data to directly disentangle different mechanisms of the rebound effect. Yet, through building characteristics data, we find heterogeneous effects that may drive the rebound effects on HVAC and dishwasher upgrade programs. Energy savings are not delivered in older houses that upgrade HVAC.

Our study highlights the potential for improving EE program effectiveness by choosing the appropriate subsidy for different products. The results also indicate that policymakers should reconsider the allocation of funds not simply based on engineering models, but directed toward programs that generate larger measured impacts. For example, lighting programs, which are promoted as an effective energy efficiency upgrade worldwide, may only have minor effects on energy savings in California. Finally, if some programs do not deliver expected energy savings, policymakers should reconsider their policy goals and the targeted population vs. actual participants in these programs. Some upgrade programs, such as HVAC, whole house retrofit, and dishwasher incentives, may generate side benefits such as home comfort and convenience, but progress toward the overall environmental target of reducing electricity consumption may be questionable.

It is worth noting that one limitation of this analysis results from the categorization of some product types, for example, HVAC upgrades do not clearly distinguish between AC units and heating systems. The inability to distinguish two separate energy delivery systems limits our ability to evaluate the effectiveness of HVAC measures. Therefore, for future evaluation purposes, we recommend creating a more expressive classification field in the program data. This classification could be similar to the "technology category" as defined in the Building Energy Data Exchange Specification (BEDES), a dictionary of terms developed by the U.S. Department of Energy for stakeholders making important energy investment decisions. We recommend that within this classification scheme, categories such as HVAC, which represent a broad collection of energy services (i.e. heating, ventilation, cooling), should be broken down into individual categories. Each of these categories could then be separately analyzed for performance/effectiveness. Furthermore, while electricity is a major energy source for residential homes in the US, natural gas is also a significant source of energy consumption. In future work, it will be useful to evaluate the impacts of subsidy programs on natural gas consumption.

Our analysis draws attention to some important avenues for future research. As our focus of this study is on the savings impact of various EE upgrade incentives, the low take-up of energy efficiency programs is still puzzling. We need to better understand how different program designs affect adoption rates. For example, pool pump programs may have great savings potential, but low adoption potential in a low-income neighborhood, while HVAC and whole house retrofit programs may exhibit greater adoption potential for low-income families, and generate larger side benefits, such as home comfort. To better understand this distributional effect, we would need to conduct household-level surveys and incorporate these results into electricity consumption data. For example, there is room for future research to examine the optimal cost-sharing amount for low income households as compared to high income household.

Furthermore, we find that some incentive programs for energy efficiency upgrades are relatively more cost-effective than behavioral programs, and may generate more persistent long-term saving results. While we compare our cost-effectiveness figures with the literature through similar calculations, prospective work is needed to directly compare energy efficiency programs and other types of behavioral programs through large-scale randomized controlled trials.

References

Adan, H., and F. Fuerst (2016). Do Energy Efficiency Measures Really Reduce Household Energy Consumption? A Difference-in-Difference Analysis. *Energy Efficiency*, 9(5), 1207-1219.

Allcott, H., and M. Greenstone (2012). Is There an Energy Efficiency Gap? *The Journal of Economic Perspectives*, 26 (1), 3–28.

Allcott, H., and M. Greenstone (2017). Measuring the Welfare Effects of Residential Energy Efficiency Programs. Unpublished Manuscript.

Allcott, H. and S. Mullainathan (2010). Behavior and energy policy. *Science*, 327(5970), 1204-1205.

Alix-Garcia, J. M., K. R. Sims, and P. Yañez-Pagans (2015). Only one tree from each seed? Environmental effectiveness and poverty alleviation in Mexico's Payments for Ecosystem Services Program. *American Economic Journal: Economic Policy*, 7(4), 1-40.

Arimura, T. H., S. Li, R. G. Newell, and K. Palmer (2011). Cost-effectiveness of Electricity Energy Efficiency Programs. Unpublished Manuscript.

Arkes, H. R. and C. Blumer (1985). The Psychology of Sunk Cost. *Organizational Behavior and Human Decision Processes*, 35 (1), 124–140.

Asensio, O. I. and M. A. Delmas (2016). The dynamics of behavior change: Evidence from energy conservation. *Journal of Economic Behavior & Organization*, 126, 196-212.

Asensio, O. I. and M. A. Delmas (2017). The Effectiveness of US Energy Efficiency Building Labels. *Nature Energy*, 2, 17033.

Auffhammer, M., C. Blumstein, and M. Fowlie (2008). Demand Side Management and Energy Efficiency Revisited. *Energy Journal*, 29(3), 91–104.

Barbose, G., C. Goldman, I. Hoffman, and M. Billingsley (2013). The Future of Utility Customer-Funded Energy Efficiency Programs in the United States: Projected Spending and Savings to 2025.

Boomhower, Judson, and Lucas W. Davis (2014). A Credible Approach for Measuring Inframarginal Participation in Energy Efficiency Programs. *Journal of Public Economics*, 113, 67–79.

Boomhower, J. P., and L. W. Davis (2017). Do Energy Efficiency Investments Deliver at the Right Time? Unpublished Manuscript.

Borenstein, S. (2005). The Long-Run Efficiency of Real-Time Electricity pricing. *The Energy Journal*, 93–116.

Chen, V. L., M. A. Delmas, W. J. Kaiser, and S. L. Locke (2015). What Can We Learn from High-frequency Appliance-level Energy Metering? Results from a Field Experiment. *Energy Policy*, 77, 164-175.

Qiu, Y., and A. Patwardhan (2018). Big Data and Residential Energy Efficiency Evaluation. *Current Sustainable/Renewable Energy Reports*, 1-9.

California Public Utilities Commission (2011). California Energy Efficiency Strategic Plan: January 2011 Update. San Francisco, CA: CPUC.

California Public Utilities Commission (2015). 2010–2012 Energy Efficiency Annual Progress Evaluation Report. San Francisco, CA: CPUC.

California Public Utilities Commission (2016). Regulating Energy Efficiency—A Primer on the CPUC's Energy Efficiency Programs. San Francisco, CA: CPUC.

Consortium for Energy Efficiency (CEE) (2008). Energy Efficiency Programs: A $3.7Billion U.S. and Canadian Industry. 2007 Report.

Consortium for Energy Efficiency (CEE) (2012). State of the Efficiency Program Industry: Budgets, Expenditures and Impacts 2011.

Cattaneo, M. D. (2010). Efficient Semiparametric Estimation of Multi-Valued Treatment Effects under Ignorability. *Journal of Econometrics*, 155 (2), 138–154.

Davis, L. W., A. Fuchs, and P. J. Gertler (2014). Cash for Coolers: Evaluating a Large-Scale Appliance Replacement Program in Mexico. *American Economic Journal: Economic Policy*, 6 (4), 207–38.

Davis, L. W., and P. J. Gertler (2015). Contribution of Air Conditioning Adoption to Future Energy Use under Global Warming. *Proceedings of the National Academy of Sciences*, 112 (19), 5962–5967.

DNV.GL (2014). HVAC Impact Evaluation FINAL Report WO32 HVAC – Volume 1: Report.

Fowlie, M., M. Greenstone, and C. Wolfram (2015). Do Energy Efficiency Investments Deliver? Evidence from the Weatherization Assistance Program. Unpublished Manuscript.

Friedman, D., K. Pommerenke, R. Lukose, G. Milam, and B. A. Huberman (2007). Searching for the Sunk Cost Fallacy. *Experimental Economics*, 10 (1), 79–104.

Gillingham, K. and K. Palmer (2014). Bridging the Energy Efficiency Gap: Policy Insights from Economic Theory and Empirical Evidence. *Review of Environmental Economics and Policy*, 8 (1), 18–38.

Gillingham, K., D. Rapson, and G. Wagner (2016). The Rebound Effect and Energy Efficiency Policy. *Review of Environmental Economics and Policy*, 10(1), 68-88.

Giraudet, L. G., and S. Houde (2015). Double Moral Hazard and the Energy Efficiency Gap. Unpublished Manuscript.

Hong, S. H., T. Oreszczyn, I. Ridley, and Warm Front Study Group. (2006). The Impact of Energy Efficient Refurbishment on the Space Heating Fuel Consumption in English Dwellings. *Energy and Buildings*, 38(10), 1171-1181.

Houde, S and J. E. Aldy (2017). Consumers' Response to State Energy Efficient Appliance Rebate Programs. American Economic Journal: Economic Policy, 9 (4), 227-255

Imbens, G. W. and J. M. Wooldridge (2009). Recent Developments in the Econometrics of Program Evaluation. *Journal of Economic Literature*, 47 (1), 5–86.

Joskow, P. L., and D. B. Marron (1992). What Does a Negawatt Really Cost? Evidence from Utility Conservation Programs. *The Energy Journal*, 13 (4), 41–74.

Loughran, D. S., and J. Kulick (2004). Demand-Side Management and Energy Efficiency in the United States. *The Energy Journal*, 19–43.

McKinsey and Company (2009). Unlocking Energy Efficiency in the U.S. Economy. Technical report, McKinsey Global Energy and Materials.

Glazerman, S, D.M. Levy, and D. Myers (2003) Nonexperimental versus experimental estimates of earnings impacts. *The Annals of the American Academy of Political and Social Science* 589 (1), 63–93.

Guimaraes, P. and P. Portugal (2010). A Simple Feasible Alternative Procedure to Estimate Models with High-Dimensional Fixed Effects. *Stata Journal*, 10 (4), 628–649.

Rubin, D. B. (2001). Using Propensity Scores to Help Design Observational Studies: Application to the Tobacco Litigation. *Health Services and Outcomes Research Methodology*, 2 (3–4), 169–88.

Rubin, D. B. (2007). The design versus the analysis of observational studies for causal effects: parallels with the design of randomized trials. *Statistics in medicine*, 26 (1), 20–36.

Rosenbaum, P. R., and D. B. Rubin (1985). Constructing a Control Group Using Multivariate Matched Sampling Methods That Incorporate the Propensity Score. American Statistician, 39 (1), 33–38.

Scheer, J, M. Clancy, and S. N. Hógáin (2013). Quantification of Energy Savings from Ireland's Home Energy Saving Scheme: An Ex Post Billing Analysis. *Energy Efficiency*, 6 (1), 35-48.

Stuart, E. A. (2010). Matching Methods for Causal Inference: A Review and a Look Forward. *Statistical Science*, 25 (1), 1–21.

Thaler, R. (1980). Toward a Positive Theory of Consumer Choice. *Journal of Economic Behavior and Organization*, 1 (1), 39–60.

The Guardian (2009, May 23). America's New Green Guru Sparks Anger over Climate Change U-turns. *The Guardian*, Retrieved from

Trotta, G. (2018). The Determinants of Energy Efficient Retrofit Investments in the English Residential Sector. *Energy Policy*, 120, 175-182.Zivin, J. G., and K. Novan (2016). Upgrading Efficiency and Behavior: Electricity Savings from Residential Weatherization Programs. *The Energy Journal*, 37(4), 1–23.