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

Yating Chuang, Magali A. Delmas, Stephanie Pincetl

Abstract: We analyze multiple subsidy programs for residential energy efficiency upgrades from 2010 to 2015 using electricity billing records of more than 11 million households in Southern California. We find that adopting these upgrades reduces overall electricity usage by 4%. However, there are significant differences in savings between upgrades. Pool pump and refrigeration upgrades generate the largest savings (13% and 6%, respectively). Lighting and HVAC retrofits generate the smallest savings (less than 1%). Some upgrades lead to concerns of rebound effects, such as dishwasher and clothes washer upgrades, and building envelope upgrades. Program impact varies by time of the year and building type. Furthermore, we find that energy savings are inconsistent with the engineering estimates. These results indicate that policy makers should consider the allocation of program funding not simply based on engineering projections but also based on measured electricity consumption such as those described in this study.

JEL Codes: L68, L94, Q41

Keywords: energy efficiency, program evaluation, residential buildings

ENERGY EFFICIENCY (EE) is one of the main policy tools for addressing climate change. Former US 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" (*Guardian* 2009). EE subsidies are politically attractive because of their ability

Yating Chuang is in the Department of Economics, National Taipei University (yating@gm.ntpu.edu.tw). Magali A. Delmas is at the University of California Los Angeles Institute of the Environment and Sustainability and Anderson School of Management (delmas@ucla.edu). Stephanie Pincetl is at the University of California Los Angeles Institute of the Environment and Sustainability (spincetl@ioes.ucla.edu). We thank project manager Felicia Federico for her essential
Dataverse data: https://doi.org/10.7910/DVN/OZTRXJ

Received January 14, 2020; Accepted December 10, 2021; Published online April 18, 2022.

Journal of the Association of Environmental and Resource Economists, volume 9, number 4, July 2022.

https://doi.org/10.1086/718529

to reduce energy usage as well as save money for consumers and governments. As a result, governments around the world are developing policies to encourage energy efficiency. California, for example, set an ambitious target of reducing its greenhouse gas emissions to 40% below the 1990 level by 2030 (SB-32 2016). One critical pathway identified to achieve this goal is to subsidize energy efficiency upgrades (CPUC 2016). As such, the state of California spends about $1 billion annually on residential energy efficiency upgrade programs.

There is, however, inconclusive empirical evidence about the effectiveness of these programs. Most claims regarding savings resulting 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 model of 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 designs suggests that these programs seldom deliver the savings predicted by engineering estimates (Davis et al. 2014; Graff Zivin and Novan 2016; Allcott and Greenstone 2017; Liang et al. 2017; Fowlie et al. 2018).

Scholars have argued that there is a need for additional empirical research to identify differing program impacts as a result of program design, household type, and building characteristics (Allcott and Greenstone 2012). However, the difficulty of accessing energy consumption data and the low take-up rates of EE programs hamper such empirical analyses. Widespread privacy laws limit access to high-resolution energy data by building, and large-scale evaluations using credible quasi-experimental design are still rare (Mathew et al. 2015; Pincetl et al. 2016). In addition, the low take-up of EE programs is a common challenge for undertaking robust analyses. For example, Fowlie et al. (2018) conducted the nation's largest randomized experiment to encourage households to adopt a free energy retrofit program in Michigan. The take-up rate was only 6% despite

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 Institute of the Environment and Sustainability (IoES) research seminar, the nineteenth annual CU Environmental and Resource Economics Workshop, the USC tenth International Workshop on Empirical Methods in Energy Economics at the University of Southern California, 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 policy makers from Los Angeles County. We are grateful to editor David Popp and three anonymous reviewers for their detailed and constructive comments and suggestions.

ample resources spent for recruitment. Overall, 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 Patwardhan 2018).

Against this background, we have compiled, to our knowledge, the most comprehensive micro data set to evaluate the effectiveness of various residential EE upgrades. First, Southern California Edison (SCE) granted us access to information on all their residential EE upgrade programs from 2010 to 2015, including records of all financial incentives claimed by SCE customers as part of these programs. Second, we obtained meter-based monthly electricity consumption data for approximately 11 million households in SCE's territory from 2010 to 2014. Data were accessed through the LA Energy Atlas, a relational database that tracks energy usage across LA County as well as building characteristics such as vintage and square footage (Pincetl and LA Energy Atlas Development Team 2015). The data allow us to address the following research questions: (1) How effective are residential EE upgrades (appliance and equipment upgrades) at saving energy? (2) Does effectiveness vary across income groups and building characteristics? (3) What is the difference between engineering estimates of potential energy savings and actual measured energy savings among different upgrades? To evaluate the savings of different EE upgrades, we specifically focus on electricity savings, as this is a very important claim of benefits from energy efficiency programs (Energy Commission and Public Utilities Commission 2005). This allows us to compare our results with the existing literature on EE programs (Allcott and Mullainathan 2010; Davis et al. 2014; Asensio and Delmas 2017).

Our study makes two contributions. First, our rich program data allow us to refine our understanding of the effectiveness of different types of EE upgrades. The paper adds to the literature using large-scale micro data to evaluate energy efficiency programs. For example, two prominent studies focus on residential energy efficiency programs and low-income weatherization programs in Wisconsin and Michigan (Allcott and Greenstone 2017; Fowlie et al. 2018). 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),[1] and/or a single demographic (e.g., low income), it is unclear whether the ineffectiveness occurs because of the program or because of household demographics. We believe that we provide the first study comparing numerous EE upgrades using large-scale micro data consumption (e.g., more than 11 million households' utility billing records) while including building information. Our extensive coverage of program data allows us not only to compare different types

1. Allcott and Greenstone (2017) focus on two energy efficiency programs in Wisconsin, which are part of the national Better Building Neighborhood Program (BBNP). The two programs were targeted at retrofitting residential buildings. The energy consultant first gives 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.

of subsidies, both in terms of technology supported and incentive schemes (such as free giveaways or cash rebates) but also to control for overlapping EE upgrade impacts (owing to the fact that EE participants may apply to multiple programs), therefore avoiding overestimation that may result from these overlaps.

Our second contribution is to examine heterogeneous effects by connecting EE upgrades with building characteristics, such as vintage and square footage. This differs from previous studies that typically only link account-level electricity information to census sociodemographics, without including building-level information. Observing heterogeneous effects of product upgrades is critical to refining funding allocation and program design. For example, the realized savings of a heating, ventilation, and air conditioning (HVAC) unit upgrade can be greater in a new and well-insulated building than in an older building which is not well insulated. Few studies contain information on building characteristics, and those that do, include insufficient numbers of observations to conduct a well-powered analysis of heterogeneity. Thus, our large-scale data analysis contributes to the literature by examining heterogeneous effects with a higher degree of accuracy. Table A1 (tables A1–A20 are available online) shows how our approach contributes to the literature by listing the most recent academic research using micro data to evaluate residential energy efficiency upgrade programs.[2] This information has been crosschecked with the most recent review paper from experts in this field (Qiu and Patwardhan 2018).[3]

To evaluate the savings of participation in different EE upgrades (monthly kilowatt-hours [kWh]), we follow the econometric strategy of prominent research in EE program evaluation (Davis et al. 2014; Liang et al. 2017).[4] Since participation in EE upgrade programs is voluntary, we implement matching procedures to address self-selection bias in program participation and then conduct post-matching difference-in-difference strategy regressions with various fixed effects (household-month and time). 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 covariate matching with the Mahalanobis metric on a pool of more than 11 million households, based on building characteristics geocoded from the assessors' database and census block group

2. We have compared our study with a substantial amount of relevant literature (Hong et al. 2006; Scheer et al. 2013; Boomhower and Davis 2014; Alberini and Towe 2015; Davis and Gertler 2015; Adan and Fuerst 2016; Alberini et al. 2016; Aydin et al. 2017; Giraudet et al. 2018; Novan and Smith 2018; Davis et al. 2020; Liddle et al. 2020; Adekanye et al. 2021).

3. However, we do not include information and behavioral programs and other macro or utility-level analyses as they are very different from our study. Gillingham et al. (2018) provides a more comprehensive review on energy efficiency. For energy conservation and information programs review, see Delmas et al. (2013).

4. Note that 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.

data.[5] Second, we use panel regression models to estimate the average savings after matching. We exploit the monthly variation of program participation and use household-month and time fixed effects to rule out time-invariant unobservable confounding factors. We also control for time-varying trends at the household level and conduct alternative quasi-experiment design among only participating households. The results are robust to various model specifications, such as matching on different sets of observables, matching with and without replacement, matching on location, matching on location and past electricity consumption, and so forth.

It is important to note that our data do not allow us to rule out the fact that some households in the sample might be those who needed a new appliance. In such cases the subsidies might be encouraging them to purchase a more efficient model than they would otherwise. Without a clean no-policy scenario data or a randomized control study, we may overestimate the savings impacts. Nonetheless, we can compare the impact of these energy upgrades on energy consumption.

Our results show that EE product upgrades from 2010 to 2015 reduce overall electricity usage by 4%. However, the energy savings vary significantly by product type. While pool pumps and refrigeration upgrades are associated with significant energy savings (12%–13% and 6%, respectively), HVAC retrofits generate insignificant overall savings. Other types of upgrades, such as clothes washers and dishwashers, and building envelope are associated with zero savings or even some increases in overall consumption.[6]

In addition, we find that energy savings are considerably inconsistent with the engineering estimates computed by Southern California Edison, the investor-owned utility that covers most of Southern California and that is responsible for the EE incentive programs. For example, lighting upgrades achieve only 7% of the engineering estimates;[7] and whole house retrofits achieve just 18% of the engineering estimates, while

5. We use exact match on building type (single family, multifamily, condo, etc.), vintage bins, climate zone, square footage percentile (only for single and multifamily housing), also fuzzy nearest distance matching on census block group–level variables, such as median income, density, poverty, white, black, Asian, Hispanic, education, age, percentage of ownership, percentage of occupancy rate, and also variables at the account level, including, whether homeowner or not, geographic location—x-y coordinates, whether the household is registered under CARE/FERA energy discount programs for low-income households, and whether the household is identified as having a pool or not.

6. The results for whole house retrofits are less conclusive. In most specifications, whole house retrofits are associated with increased savings, but in some specifications, they are associated with zero to 3 percentage point savings. In the falsification test, it appears that program participants have already experienced an increased upward trend in electricity consumption, creating a concern of identification assumption.

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

other upgrades such as those for dishwashers and clothes washers generate zero savings or even increased energy consumption. However, pool pump upgrades 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 the heterogeneity of upgrade effectiveness, we find that the level of the financial incentives matters. For example, we find larger savings for some of the products with cost sharing (e.g., lighting rebate programs) as compared to 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 upgrades as the savings may largely depend on recipients' usage behaviors. Indeed, it is possible that free light bulbs are not installed by households, in contrast with light bulbs purchased through rebates.[8]

The paper proceeds as follows. Section 1 provides background on energy efficiency programs. Section 2 presents our data and empirical strategy. Section 3 presents our main results. Section 4 details heterogeneous estimates. Section 5 presents robustness checks. Section 6 provides a concluding discussion.

1. PROGRAM BACKGROUND

Our study focuses on all the customer-incentive-based EE programs administered from 2010 to 2015 by Southern California Edison (SCE)—one of the largest utilities in the United States, providing electricity to more than 14 million people. The weather in much of this region includes warm, dry summers and mild winters. Due to the lack of exceptionally cold weather, about 14% of California homes are not heated (EIA 2009). This weather pattern may affect energy savings of more weather-dependent energy efficiency products such as HVAC, whole house retrofits, and building envelope.[9]

During the study period, SCE provided 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 during the program implementation period, SCE customers needed to upgrade to a more energy efficient product in their homes and then apply for the rebates by mail or online. For example, to apply for an appliance upgrade (e.g., HVAC, pool pump, refrigerator) single-family homeowners could apply for rebates through the Home Energy Efficiency Rebate (HEER) Program—the largest residential program based on expenditure (CPUC 2015)—and multifamily owners could apply through the Multifamily Energy Efficiency Rebate

8. However, it is worth noting that this difference between cost sharing and free delivery is not evident among other programs because of data limitation.

9. We do not account for natural gas consumption, which is an important component when evaluating weather-dependent energy efficiency upgrades. This limitation may be less of a problem in our study because the winter is relatively mild in our study region. However, one needs to be careful when generalizing our results to other regions.

Program. 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.

These EE programs include nine types of products based on SCE's categorization: appliance upgrade (i.e., dishwasher and clothes washer), consumer electronics, HVAC retrofit, lighting, pool pumps, refrigeration (i.e., refrigerator and freezer), audits, whole house retrofit, building envelopes, and others.[10] Note that households may claim multiple incentives to upgrade their homes through various programs. Our estimates, therefore, provide the average monthly impact (e.g., electricity reduction) of a typical household that has claimed upgrade rebates within our study period. To be consistent with SCE's program approach, we focus our analysis on product categories. Our category selection is based on those used by the Database of Energy Efficiency Resources (DEER). Among those listed, lighting and refrigeration have attracted the largest consumer participation. Table 1 displays the number of SCE households enrolled and the products covered in the SCE EE programs. We also present the detailed program implementation strategies in appendix B (apps. A–C are available online).

EE programs use different levels and types of financial incentives. For example, some subsidize a certain percentage of the cost of the product (i.e., rebate), while others provide the products for free. Programs can also be directed to different types of recipients: upstream and midstream incentives are given to contractors or distributors, while downstream incentives are given directly to the end users (households). A full list of the different levels and types of incentives is shown in table 2.

2. DATA AND EMPIRICAL FRAMEWORK

2.1. Data Description

We combine four data sets to understand the effectiveness of energy efficiency upgrades among residential households in Southern California. We focus on energy efficiency upgrade programs where households could claim financial support between 2010 and 2015.[11] To evaluate the impact of these EE upgrades 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

10. The term "building envelope" refers to the external elements of the building that enclose the internal space, including windows, exterior four walls, roof, and the floor above the unheated basement area.

11. Program-related information (e.g., installation date, rebate amount, etc.) was recorded. However, we do not analyze 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.

Table 1. Energy Efficiency Technologies

EE Program	Appliance	Consumer Electronics	HVAC	Lighting	Pool Pump	Refrigeration	Water Heating/ Savings Kit	Audit	Whole House Retrofits	Building Envelope	Other	Total	Program Cycle
Desert Cities Energy Leader Partnership	0	0	3	15	342	0	0	0	0	0	59	419	2010–12
Palm Desert Demonstration Partnership	0	0	1,840	1,508	688	73	0	0	0	0	1,779	5,888	2010–12
Home Energy Efficiency Survey Program	0	0	0	127,016	0	0	125,565	0	0	0	283,008	535,589	2010–12
Home Energy Efficiency Rebate Program	6	37	5,431	0	13,019	99,693	314	0	0	0	14	118,514	2010–12
Appliance Recycling Program	0	0	0	0	0	149,282	0	0	0	0	0	149,282	2010–12
Business and Consumer Electronics Program	0	107	0	0	0	0	0	0	0	0	0	107	2010–12
Multifamily Energy Efficiency Rebate Program	3	0	89	117,859	0	1,040	52	0	0	0	784	119,827	2010–12
Whole House Prescriptive Program	0	0	0	0	0	0	0	0	0	0	21	21	2010–12
California Advanced Homes	2,440	0	171	1	0	1,680	0	0	0	0	3,760	8,052	2010–12
Energy Star Residential Quality Installation Program	0	0	4,218	0	0	0	0	0	0	0	0	4,218	2010–12
Residential Quality Mainte-nance and Commercial Quality Maintenance Development	0	0	1,776	0	0	0	0	0	0	0	0	1,776	2010–12

Comprehensive Mobile Home	0	0	2,820	8,002	0	1	0	0	0	0	0	10,823	2010–12
Comprehensive Home Performance	0	0	0	0	0	0	0	0	0	0	518	518	2010–12
Coin Operated Laundry Program	11	0	0	184	0	0	0	0	0	0	0	195	2010–12
California Statewide Program for Residential Energy Efficiency	13,561	364	8,161	71,700	18,915	108,232	3,597	219,478	5,893	1,648	0	451,549	2013–15
Comprehensive Home Performance	0	0	0	0	0	0	0	0	0	0	518	518	2013–15
Comprehensive Manufac-tured Homes	0	0	9,512	11,069	12	1	0	0	0	0	0	20,594	2013–15
Energy Leader Partnership Program	0	0	1,786	845	1,104	0	0	75	0	4	1,533	5,347	2013–15
Lighting Program	0	0	0	5,433	0	0	0	0	0	0	0	5,433	2013–15
New Construction Program	2,552	0	171	1	0	1,756	0	0	3,199	0	3,759	11,438	2013–15
Residential and Commercial HVAC Program	0	0	4,959	0	0	0	0	0	0	0	0	4,959	2013–15
Residential Energy Efficiency Program	10	144	5,511	249,487	13,015	248,084	125,936	0	0	13	283,803	926,003	2013–15
SCG Co-fund	12,461	0	856	0	0	0	0	0	0	2,989	0	16,306	2013–15
Statewide Commercial Energy Efficiency Program	0	0	225	0	0	0	0	0	0	0	0	225	2013–15
Total	31,044	652	47,529	593,120	47,095	609,842	255,464	219,553	9,092	4,654	579,556		

Note. The number represents the total count of unique participating households.

Table 2. Energy Efficiency Upgrade by Type of Subsidies

Product	Upstream/ Midstream	Financial Incentives (Medium)	Direct Install	Free/Give-away (Large)	No. of Subsidy Claims
Appliance	118	30,082	0	0	30,200
Consumer electronics	73,077	0	364	0	73,441
HVAC	381	25,597	59,664	0	85,642
Lighting	66,343	337,857	531,422	448,752	1,384,374
Pool pump	6,924	35,482	480	0	42,886
Refrigeration	76	464,673	0	0	464,749
Water heating/ savings kit	0	1,106	8,165	260,894	270,165
Audits	0	224,361	0	76	224,437
Whole house retrofit	6,872	2,459	0	0	9,331
Building envelope	0	5,526	0	0	5,526
Other	0	5,896	0	293,012	298,908
Total	153,791	1,133,039	600,096	1,002,734	2,889,659

County (Pincetl and LA Energy Atlas Development Team 2015).[12] Key data for EE program participation are from both the California Public Utilities Commission (CPUC) and Southern California Edison (SCE), the regional electricity utility. Building-level characteristics come from the Los Angeles County Assessor's property data set. Sociodemographic information is from the census database. (See table A2 for more details regarding the source and coverage of our data.)

2.1.1. EE Program Data

Program participation data identify residential energy efficiency programs implemented in SCE service territories during 2010–15. Note that 2015 electricity consumption data are unavailable to us, so to assess product upgrade impact on electricity savings, we analyze information only from 2010 to 2014. However, program information from 2015 is useful to prevent us from selecting any future EE participants into our matched control group.

Each time a household claims a rebate or direct financial support to upgrade their home, there is a record that documents important information, such as the installation date, rebate amount, type of product, and predicted energy savings. We use the installation date to generate an EE upgrade participation variable reflecting the starting

12. See www.energyatlas.ucla.edu.

month of the upgrade for each household and each product category. All information tracked after the installation date *t* is considered part of the "treatment period." Note that since we have a short five-year window of data, it is fair to assume that all upgrades remain within their effective lifespan throughout our study. The other important variable is the predicted savings. For each claimed incentive for each type of product, we can calculate the estimated savings based on SCE's internal methodology.

The EE program participation data cover 191 cities (see fig. A1; figs. A1–A5 are available online).[13] Participation comparisons across income categories show higher participation rates from the top income quartile (see table A3). The average uptake rate is 8%, while it is 11.6% within the highest income quartile and 5.5% within the lowest income quartile. Orange County has higher EE adoption rates (12%) than other counties, while the adoption rates of Los Angeles County and Imperial County are lower than average (around 6%). Given the vast variation, we opted for a credible quasi-experimental design to mitigate self-selection bias and best evaluate EE upgrade effectiveness.

2.1.2. Electricity Usage Data

Account-level electricity billing data are available from January 2010 through December 2014 across the SCE service territory. Our data contain more than 11 million unique accounts with monthly electricity usage data from 2010 to 2014. The panel data are unbalanced because households may move in and out of a given building, or even area, during the study period, which means that not all households have a complete record running throughout the entirety of the study period.

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 the same utility account, we generate a different ID. This approach accounts for the potential that the same household may consume electricity differently in different homes due to building characteristics such as vintage and square footage. Also, most upgrades are attached to the building and will not be carried over once households move to a new building. So treated households are no longer considered treated once they move.

2.1.3. Building Characteristics Data

Building characteristics used in this analysis include (1) use type (i.e., single-family housing, multifamily housing, condominium, etc.), (2) square footage, (3) building vintage, (4) climate zone, (5) energy discount program participation (for low-income households), (6) whether the house is the owner's primary residence, (7) whether the

---

13. 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.

household is registered under low-income discount programs, and (8) whether the household is identified as having a swimming pool. The descriptive analysis of the key variables is shown in table A2. For LA County, building use type and ownership information is sourced from the 2016 county assessor office's parcel database, which 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 multistage geocoding process developed by 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.

2.1.4. Sociodemographic Data

We use US census data to obtain sociodemographic information. The census information is taken from the American Communities Survey (ACS) 2006–10 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.

2.2. Empirical Strategy

We can identify the savings for each product upgrade by comparing the change of electricity consumption over time between participants and nonparticipants under several identification assumptions (see details in the identification section). The challenge in evaluation is that many factors other than energy use may influence whether households participate in an EE upgrade. In order to rule out factors that may confound the EE upgrade, we first find a set of households that were similar to our EE product upgrade participants but who had never claimed any EE upgrade subsidy. We use a matching method to find this comparison group and based our matching criteria on variables cited to be important in previous studies such as location, building characteristics, and demographic variables. As recommended by Rubin (2007), we do not use any outcome variable, such as energy consumption, to construct our control group so as to retain the objectivity of our design.[14] The literature shows that regression models may perform poorly with small covariate overlap (Dehejia and Wahba 1999, 2002; Glazerman et al. 2003; Alix-Garcia et al. 2015), so this pre-match method aimed to 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). As expected, our covariates are more balanced after matching (see fig. A2 and matching covariates test results in table 3).

---

14. Nevertheless, we tried matching on location and matching on location and past consumption pattern as in Davis et al. (2014), and the results are robust.

We match households that participate in EE upgrades (treated households) in any year between 2010 and 2015 with households that have never participated in any EE upgrades using Mahalanobis Distance Matching with replacement.[15] This matching scheme uses the data from over 10 million nonparticipants and finds the nearest neighbors to our treated households based on the following covariates: exact matching for (1) building type (single-family housing, multifamily housing, condominium, combined residential), (2) vintage bins (built before 1950, between 1950 and 1978, between 1979 and 1990, and after 1990), (3) square footage bins, (4) climate zone, (5) fuzzy nearest distance matching for other covariates including owner-/renter-occupied,[16] (6) whether the household participated in any low-income electricity discount program (e.g., California Alternate Rates for Energy [CARE] or Family Electric Rate Assistance [FERA]), (7) whether the household had a pool,[17] and (8) census block group variables, such as median income, density, poverty, race, education, age, percentage of ownership, and percentage of occupancy rate. We also include x-y geographic coordinates to improve our location matching.[18] 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 upgrades in our study. Our advantage is to use both sociodemographic and building-level information in selecting our counterfactual group, as the existing literature uses covariates from sociodemographic information (mostly at the census block group level) or past electricity usage to find the counterfactual. Based on table A1, very few studies have building-level information because billing data do not include building characteristics. This limitation may constrain econometric strategies given that building characteristics are considered to be more important determinants in explaining energy efficiency retrofit investment than sociodemographic characteristics (Trotta 2018).

15. We also tried without replacement and with different cluster level and the results are similar.

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

17. Pool ownership information is limited and incomplete. Therefore, for those who do not have pool information, we 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 also tried matching without this pool ownership information, and the result is consistent.

18. Since we use x-y coordinates in matching, some may worry about spillover effect—nonparticipants 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 savings impacts underestimated, which means that our identified savings, 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.

Table 3. Summary Statistics: Energy Efficiency Upgrade Participants and Nonparticipants

	All (1)	Non-EE Participants (2)	EE Participants (3)	Matched Non-EE Participants (4)	Normalized Difference (3) versus (2) (5)	Normalized Difference (3) versus (4) (6)
Account level:						
Square footage	1,994	1,996	1,981	2,007	−.002	−.004
	(5,646)	(5,819)	(3,843)	(5,377)		
% homeowner also resident	32%	30%	53%	44%	.335	.146
	(.46)	(.46)	(.50)	(.50)		
% of people with pool	13%	13%	18%	14%	.099	.071
	(.34)	(.33)	(.38)	(.35)		
% of people under CARE/FERA	22%	21%	38%	34%	.281	.058
	(.41)	(.40)	(.49)	(.48)		
Census block group:						
Median income	66,229	65,428	75,487	70,148	.215	.117
	(31,595)	(31,176)	(34,782)	(34,936)		
Population density per square mile	9,798	9,958	7,944	7,969	−.183	−.011
	(8,376)	(8,463)	(7,029)	(7,424)		
Population under poverty	348	354	285	324	−.142	−.090
	(358)	(361)	(317)	(355)		
Total population: white	1,298	1,293	1,353	1,308	.047	.036
	(916)	(917)	(911)	(866)		

Total population: black	173	176	142	153	−.087	−.035
	(290)	(292)	(256)	(263)		
Total population: Asian	293	290	329	276	.065	.087
	(418)	(416)	(445)	(407)		
Total population: Hispanic	950	959	844	855	−.097	−.014
	(870)	(875)	(810)	(811)		
% of population with at least a bachelor degree	43%	42%	47%	44%	.110	.067
	(.34)	(.33)	(.33)	(.34)		
% of homeowner	54%	53%	63%	60%	.276	.117
	(.27)	(.27)	(.26)	(.27)		
Occupancy rate	93%	92%	93%	91%	.064	.162
	(.09)	(.09)	(.09)	(.13)		
Age	36.4	36.3	38.4	38.0	.172	.041
	(8.6)	(8.6)	(9.2)	(9.3)		
Observations	11,042,015	10,163,364	878,651	817,585		

Note. Standard deviations are in parentheses. Matches are found through one to one covariate matching with replacement on the Mahalanobis metric. Exact matches are required on building use type (single family, multifamily, condo, etc.), vintage bins, climate zone, square footage percentile (only for single and multifamily housing). Other matching covariates are from the census block group–level variables, such as median income, density, poverty, white, black, Asian, Hispanic, education, age, percentage of ownership, percentage of occupancy rate, and also variables at the account level, including whether homeowner or not, geographic location—x-y coordinates, whether the household is registered under CARE/FERA energy discount programs for low-income households, and whether the household is identified as having a pool or not. Normalized difference is the difference in average covariate values, divided by the square root of the sum of variances for both groups (Imbens and Wooldridge 2009). The last two columns give the sum of the normalized differences across all the covariates.

After matching, we use panel regression models controlling for household fixed effects or household-month fixed effects. This approach helps 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 upgrade participation as well as electricity consumption. The specification controls for household-month fixed effects, which helps rule out household-specific unobservable seasonal usage patterns. Finally, we control for time fixed effects for any unobservable factors that may confound the EE upgrade participation.

In the post-matching estimations, we rely on the difference-in-difference (DID) technique to identify the product upgrade effectiveness. First, we determine the difference in electricity use between EE upgrade participants and nonparticipants 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 nonparticipants (second "difference").

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

$$\ln(\text{Energy}_{imt}) = \beta \text{EE}_{imt} + \alpha_{im} + \gamma_{mt} + \varepsilon_{imt}, \tag{1}$$

where EE_{imt} is an indicator variable for identifying household, i switches from zero to one when that household joined any EE upgrade in month m year t. We use the month when households installed the product to determine their treatment status. To understand energy usage, we use $\ln(\text{Energy}_{imt})$ 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 (e.g., extremely heavy energy users in Beverly Hills), we use a natural logarithm to smooth the consumption variable at the highest end of the distribution. This method also eliminates noisy observations that have zero electricity consumption, as just 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 as well as the systematic seasonal pattern of the household's electricity consumption. These household-month fixed effects address unobservable time-invariant differences in attributes between EE upgrade participants and nonparticipants. In this context, "unobservable time-invariant differences" refer to those characteristics that may affect both EE upgrade participation and electricity usage. These include, for example, a person's environmental attitudes, household size, political views, or even personal behavioral biases.

We also include month-year time fixed effects (γ_{mt}) to control for economic or administrative shocks in each time period. As savings impacts may be overstated with changes in administrative capacities over time, γ_{mt} differences out some confounding factors such as administrative and economic shocks. It also controls for the fact that the background technology of various energy efficiency products may simply improve over time.

To better understand what type of upgrade works better, we further analyze EE upgrade effectiveness for different types of products. We estimate equation (2):

$$\ln(\text{Energy}_{imt}) = \beta_1\text{EEProduct1}_{imt} + \beta_2\text{EEProduct2}_{imt} \ldots \alpha_{im} + \gamma_{mt} + \varepsilon_{imt}, \quad (2)$$

where EEProduct1_{imt}, EEProduct2_{imt} … are indicator variables that identify when household i switches from zero to one as that household upgrades an EE product (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 $\beta_1, \beta_2 \ldots \beta_{10}$). We are interested in testing the effectiveness of different end-use products (whether $\beta_1 < 0$; $\beta_2 < 0 \ldots \beta_{10} < 0$). There are 10 primary products delivered with financial incentives, including appliances, consumer electronics, HVAC, lighting, pool pump, refrigeration, audits, whole house retrofits, building envelope, and other equipment.

2.3. Identification

Our rich data allow us to control for most of the building covariates identified in the literature as important influences on the decision of whether or not to adopt EE. In addition, the large number of fixed effects helps us control for a large number of time-variant and unobservable confounds. However, it is important to note that even with our large sample and rich heterogeneous building information, our approach may not be immune to some forms of bias. Therefore, our main contribution is to examine the heterogeneous effects of different EE product upgrades. We discuss below some of the identification challenges of the analysis and how we address them.

The first potential concern is the parallel trend assumption, which requires that in the absence of treatment, the difference between "treatment" and "control" group is constant over time. We conduct robustness checks where we control for group-specific time trend. As the recent literature recommends a move away from relying purely on traditional parallel trend pretests (Bilinski and Hatfield 2019; Freyaldenhoven et al. 2019; Roth 2019; Rambachan and Roth 2021),[19] we also conduct a falsification test and sensitivity analysis (see discussion in the robustness check section). The results of these tests are robust for the most effective upgrades—pool pump and refrigeration (see detailed discussion in the robustness check section).

Second, the decision to upgrade might depend on the type of products. For example, households may decide to upgrade their dishwasher when it is broken (exogenous timing), while they may not wait for this natural transition to take on a whole house or

19. The preexisting trends for the most effective products, pool pump and refrigerator, appear to be parallel (see fig. A2).

HVAC retrofit project (endogenous timing). In addition, households may choose to enroll in a whole house retrofitting project when they have an additional household member who moves into the home or when their employment status changes. We deal with this concern through robustness checks where we control for extra household specific time control and use lagged treatment variables (see detailed discussion in the robustness check section). Moreover, the distinguishable seasonal savings patterns (in fig. 2) for HVAC and whole house retrofit reassures us that we have captured the impact from EE upgrade rather than this other confounding household event (e.g., employment change).

Third, one might be concerned that savings impacts are not completely "additional" even when the upgrade participation timing is exogenous (e.g., broken appliances). In the data, we can only observe the product upgrades covered by the program subsidy, but we cannot observe the product upgrades in the absence of these programs. Households could have purchased the very same EE upgrade without the subsidy, resulting in overestimation of our savings effects. In this case, the savings associated with product upgrades are an upper bound, since we cannot attribute all the savings impacts to the programs. Yet if we find that EE upgrades are ineffective (which happens in many of our products), the overestimation would make our result even stronger.

Despite these limitations, our study still sheds lights on sensible comparison among different EE technologies, type of subsidies, and building characteristics. Therefore, one of our contributions is to examine the heterogeneous effects of different EE product upgrades.

3. MAIN RESULTS

3.1. Summary Statistics

Table 3 shows summary statistics of various characteristics of EE participants and nonparticipants in our total sample and matched sample. EE Participants constitute 8% of the total sample.[20] They are more likely to own their homes, live in newer buildings, and reside in areas with lower population densities and higher incomes. 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.[21] This indicates that participation correlates with greater access to resources and further justifies our matching method to improve covariates overlap.

To test covariate balance, we calculate the normalized difference in means between EE participants and nonparticipants in the whole sample, and between EE participants and matched nonparticipants (see table 3 and fig. A2). The normalized difference in

20. The adoption rate for any energy efficiency upgrade during our five study years is about 8% (see also table A3).

21. All of these numbers are purely summary statistics without controlling for other covariates.

means represents 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. We calculate the standardized difference of various attributes in order to compare whether our matching method improves the similarity of the control and treatment group. This is the most common way to diagnose covariate balance (Rosenbaum and Rubin 1985; Stuart 2010).

Our matching improves covariate balance in almost every aspect (see fig. A2). Most covariates have a smaller normalized difference in means except for 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 smaller than 0.25 standard deviations—the suggested rule of thumb in the literature (Rubin 2001; Imbens and Wooldridge 2009).

3.2. Overall Treatment Effects

Our results show that residential EE upgrade incentives reduce overall electricity usage by 4%, or 26 kWh per month (see the bottom row in table 4). This is equivalent to 311 kWh of annual savings. Using this figure for 8% participating households of approximately 15 million SCE customers, we estimate that overall upgrade effectiveness is equivalent to 75 gigawatt hours in annual savings ($311 \cdot 15{,}000{,}000 \cdot 8\,\%\,/5 = 74{,}640{,}000$ kWh). This electricity reduction is equivalent to preventing 52,770 tons of carbon dioxide emissions annually.[22] Of course, these numbers are approximate as they are based on back-of-the-envelope calculations and only on the financial incentivized upgrades we evaluated. Nevertheless, they provide helpful estimates to begin to understand the impacts of EE upgrades in California.

3.3. Individual Product Effects

Beyond the overall positive results, we also examine which type of upgrade delivers the most energy savings. Table 4 presents the coefficient estimates for various EE upgrades based on equation (2). All the specifications are based on the matched sample for estimating the average treatment effect on the treated. Column 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 1–3 are clustered at the household level to control for serial correlation. In column 4, we cluster the standard errors at the building level to control for spatial and serial correlation. As standard

22. Based on the EPA's 2017 data, the emission factor in the United States is 7.07×10^{-4} metric tons CO_2/kWh. The information is retrieved from https://www.epa.gov/energy/greenhouse-gases-equivalencies-calculator-calculations-and-references.

Table 4. Impact of Energy Efficiency Upgrade on Electricity Usage

	2010–14 log(usage)			
Variables	(1)	(2)	(3)	(4)
Appliance	.0370***	.0350***	.0255***	.0255***
	(.0027)	(.0030)	(.0032)	(.0043)
Consumer electronics	−.0017	.0190	−.0108	−.0108
	(.0205)	(.0250)	(.0284)	(.0285)
HVAC	.0054**	.0131***	−.0030	−.0030
	(.0021)	(.0029)	(.0026)	(.0029)
Lighting	−.0038***	−.0079***	−.0042***	−.0042
	(.0012)	(.0013)	(.0015)	(.0026)
Pool pump	−.1186***	−.1222***	−.1266***	−.1266***
	(.0019)	(.0022)	(.0023)	(.0023)
Refrigeration	−.0639***	−.0684***	−.0617***	−.0617***
	(.0007)	(.0008)	(.0009)	(.0009)
Audits	−.0207***	−.0224***	−.0268***	−.0268***
	(.0008)	(.0008)	(.0010)	(.0010)
Whole house retrofit	.0275***	−.0086	.0147**	.0147**
	(.0052)	(.0067)	(.0065)	(.0065)
Building envelope	.0242***	.0238***	.0197***	.0197**
	(.0045)	(.0046)	(.0054)	(.0085)
Household and county-year-month fixed effect	Yes	No	No	No
Household and city-time fixed effect	No	Yes	No	No
Household-month fixed effects	No	No	Yes	Yes
Month-year fixed effects	No	No	Yes	Yes
Observations	51,441,373	42,863,329	51,441,373	51,441,373
R-squared	.8038	.8206	.9009	.9009
Overall savings:				
Effect on log(kWh per month)				−.0401***
				(.0005)
Effect on kWh per month				−25.929***
				(.525)

Note. This table reports coefficient estimates and standard errors (in parentheses) from six separate regressions—the top four are from eq. (2), and the bottom two are from eq. (1). For the top four regressions, the coefficients of interest are indicator variables for households that have participated in the EE upgrade financial incentive programs for that specific product (i.e., lighting, HVAC, etc.). For the bottom regressions estimating overall savings, the coefficients of interest are indicator variables for households who have ever participated in any EE upgrade. In almost all regressions (cols. 1–4 and the first bottom row), the dependent variable is the natural log of 2010–14 monthly electricity consumption in kilowatt-hours. Estimations are based on matching EE participants with those who have never participated in the programs in the following variables: exact match on building type (single family, multifamily, condo, etc.), vintage bin, climate zone, square footage percentile (only for single and multifamily housing), also fuzzy nearest distance matching on variables, such as median income, density, poverty, white, black, Asian, Hispanic, education, age, percentage of ownership, percentage of occupancy rate, whether homeowner is also resident or not, geographic errors clustered at the building level yield more conservative results, we use this specification in the rest of the regressions.

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 upgrades, on average, deliver 11% to almost 13% of savings; incentives on refrigeration upgrades generate 6%–7% of savings, on average. However, in the case of lighting upgrades, we find that the magnitude of savings is smaller and insignificant in the more conservative specification (col. 4) than in the less conservative specification (cols. 1 and 3).

Table 4, column 4, highlights the results based on our preferable specification—including household-month and time fixed effects and our most conservative standard error estimation. The results are illustrated in figure 1. A negative number in the figure means that EE product upgrade participants decreased electricity usage. Based on the results from all the multiple statistical models, pool pump upgrades yield the highest savings. Households participating in these upgrades, on average, reduce their energy consumption by 12%–13%. The result accounts for seasonal patterns: HVAC may be utilized more in the summer, making those upgrades attractive during certain months of the year. In this case, we may underestimate the savings 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 incentives for upgrading refrigeration (including refrigerator and freezer). Households who have new efficient refrigerators or freezers reduce their electricity consumption by 6% on average.

Lighting upgrades result in relatively small savings—0.3% to 0.7% reductions, and statistically indifferent from zero in our most conservative estimate.[23] The result could

location—x-y coordinates, whether household is registered under CARE/FERA energy discount programs for low-income households, and whether the household is identified as having a pool or not. Standard errors are clustered at the household level for the first three regressions, reported in cols. 1–3. Mean pretreatment electricity consumption is 710 kilowatt-hours per month for households who have ever participated in the energy efficiency programs. Standard errors are clustered at the building level for the last regression, reported in col. 4. Specifications using city-time fixed effects have fewer observations because households at the city boundaries are dropped. We do not report the "other" category, but to prevent from overlapping program effect we still control for "other" upgrades (including the water savings kit).

* $p < .1$.

** $p < .05$.

*** $p < .01$.

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

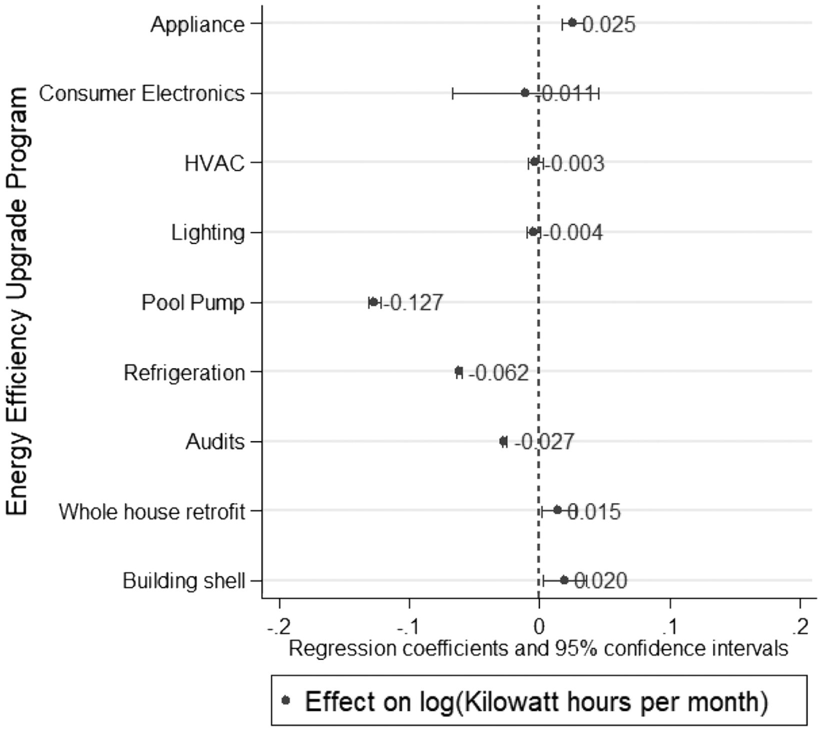

Figure 1. Energy efficiency upgrade overall regression result. Each figure plots estimated coefficients and 95th percentile confidence intervals corresponding to an indicator variable for households that have participated in the EE upgrade, one for each type of upgrade. The dependent variable in all regressions is the natural log of 2010–14 monthly electricity consumption in kilowatt-hours, and the regressions include household-month fixed effects and time fixed effects. Standard errors are clustered at the building level. Estimations are based on matching EE participants with those who have never participated in the programs in the following variables: exact match on building type (single family, multifamily, condo, etc.), vintage bin, climate zone, square footage percentile (only for single and multifamily housing), also fuzzy nearest distance matching on variables, such as median income, density, poverty, white, black, Asian, Hispanic, education, age, percentage of ownership, percentage of occupancy rate, whether homeowner is also resident or not, geographic location—x-y coordinates, whether household is registered under CARE/FERA energy discount programs for low-income households, and whether the household is identified as having a pool or not. We do not report the "other" category, but to prevent from overlapping program effect we still control for those "other" upgrades (including the water savings kit).

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 subsidies. We will not be able to directly verify the third mechanism, but we investigate the second mechanism indirectly in the next section by examining the product impact heterogeneity.

Audits, conditional on controlling for upgrading other technologies, can generate around 2%–3% savings. This impact may come from a household's behavioral change as "nudge"-style intervention can be quite successful (Allcott and Kessler 2019). We also find suggestive evidence of the behavioral change from a survey conducted by the evaluation agency. Those who participated in the SCE audits engaged in more energy savings behaviors than nonparticipants (DNV.GL 2017b).

Other upgrades do not yield large impacts in terms of electricity savings. These include appliance upgrades (mostly dishwashers and clothes washers), HVAC retrofit, whole house retrofit, and building envelope 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.[24] 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).[25] For example, after upgrading to a more energy efficient product, people may 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. We investigate this issue in the next section through an analysis of heterogeneous effects. Finally, consumer electronics upgrades yield less conclusive results across models—the lack of savings resulting from these upgrades may be due to the small sample size.

In summary, energy savings vary significantly by product type. While pool pumps and refrigeration upgrades are associated with significant energy savings (12%–13% and 6%, respectively), HVAC retrofits generate statistically insignificant savings overall.[26] Furthermore, subsidies targeting clothes washers and dishwashers, entire home retrofits, and building envelope upgrades are associated with increases in overall consumption. The results are consistent throughout the models.

---

24. 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.

25. 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.

26. This similar result was also found in previous impact studies conducted by government contracted evaluators (DNV.GL 2014a).

We also present the graphical analysis, plotting the coefficients and 95% confidence intervals for each individual product before and after the installation date. These event study figures help us visually examine the electricity consumption pattern before and after the EE upgrade. The results are presented in figure A3. The graphical analysis shows the consistency of the result—pool pump and refrigeration upgrades are effective in terms of saving energy, and the impact is quite consistent over time. The graph also confirms one of the important identification assumptions in a difference-in-difference design—the trends in the pretreatment period between control and treatment group look mostly parallel. However, assuming the pretreatment difference in trend carries out is still a strong assumption. (We provide further robustness tests in the robustness check section.)

To evaluate how the results may differ by season, we run the same model as above by month. We present our results in figure 2 (for detailed coefficients, see tables A4, A5). We find pool pump and refrigeration upgrades to be most effective, leading to savings in all seasons. This may be because pool pump and refrigeration use about the same amount of electricity throughout the year in California.[27] HVAC upgrades have a stronger seasonal effect—positive savings in the summer but none, or even negative savings, in other months. This result is consistent with the literature, which shows that HVAC upgrades deliver savings during different months of the year or even different times of day (Boomhower and Davis 2020). The result highlights the concern that the effectiveness of HVAC upgrades may be limited by the behavioral responses of the users, unlike more "passive" upgrade products like pool pump and refrigeration. Whole house retrofits exhibit seasonal savings patterns similar to HVAC upgrades but with larger confidence intervals and smaller savings in the summer months. This indicates that the effectiveness of HVAC upgrades and whole house retrofits may depend on weather and electricity usage patterns.

3.4. Comparison with Engineering Predictions

In order to quantify our estimates and compare them to the ex ante engineering estimates recorded by SCE, we estimate a regression similar to equation (2), but instead of taking the log of the dependent variable, we use the level of monthly electricity consumption in kilowatt-hours as the outcome variable. The ex ante savings predictions are based on SCE's original gross savings projection achieved by a specific physical upgrade. For each specific product upgrade, we calculate its average predicted monthly savings based on SCE's assigned life-cycle savings for that installed product. For example, most products have a life cycle of more than five years, which is longer than the time

27. The almost no seasonality result for pool pump upgrades may be perplexing. Yet we find evidence that a large portion of households operate their pool in the same manner year round in Southern California (DNV.GL 2016).

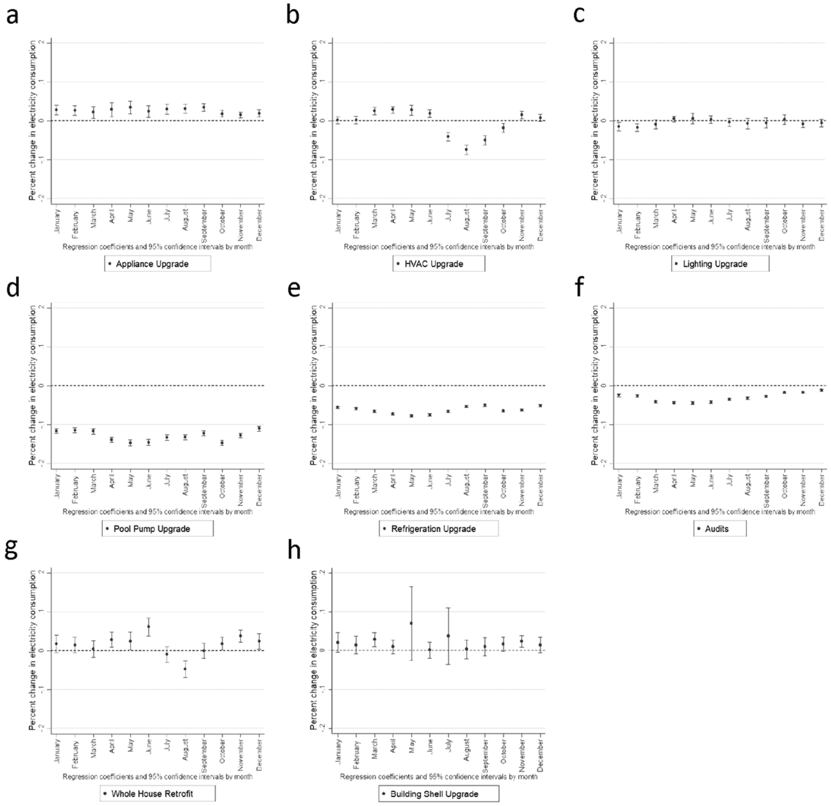

Figure 2. Percentage change in electricity consumption by product and month. Each figure plots estimated coefficients and 95th percentile confidence intervals corresponding to an indicator variable for households that have participated in the EE program, one for each month. The dependent variable in all regressions is the natural log of 2010–14 monthly electricity consumption in kilowatt-hours, and the regressions include household-month fixed effects and time fixed effects. Standard errors are clustered at the building level. Estimations are based on matching EE participants with those who have never participated in the programs in the following variables: exact match on building type (single family, multifamily, condo, etc.), vintage bin, climate zone, square footage percentile (only for single and multifamily housing), also fuzzy nearest distance matching on variables, such as median income, density, poverty, white, black, Asian, Hispanic, education, age, percentage of ownership, percentage of occupancy rate, whether homeowner is also resident or not, geographic location—x-y coordinates, whether household is registered under CARE/FERA energy discount programs for low-income households, and whether the household is identified as having a pool or not.

period of our data, so we directly divide the predicted annual savings by 12 to get monthly savings. For products that have a life cycle of less than five years, we calculate their predicted average monthly savings for five years (predicted annual savings times the life cycle of the product and divided by 60). This ex ante projection is generated based on SCE's engineering department and multiple program teams, analysts from the public and CPUC, and other industry affiliates and professionals. SCE has a so-called workpaper to determine the deemed savings based on the type of measure the customer installed, the program, and other important variables such as customers' housing, climate zone, and so forth.[28] Thus, the imposed assumptions of those ex ante savings must be approved by CPUC following specific guidelines.[29] The comparison is summarized in table 5. We also calculate the realization rate, where we divide our estimated average savings by the engineering estimates to see how much the actual savings has achieved the predicted ex ante engineering estimates. As utilities receive funding based on the predicted engineering estimates, the realization rate helps policy makers evaluate how well each type of upgrade achieves its promise.

As table 5 shows, energy savings are inconsistent with ex ante engineering estimates. For example, lighting upgrades achieve only 7% of the engineering estimates of expected savings,[30] and whole house retrofits achieve just 18%. The result of whole house retrofits seems puzzling. However, this finding is not unprecedented, as such low realization rates are also found in previous impact evaluation reports conducted by consulting companies with smaller sample sizes (12%–50% in DNV.GL [2014b]; −48% to highest 24% in DNV.GL [2017a]). Previous reports also provide some potential reasons for these findings, such as an overestimation of the ex ante savings in the retrofit planning stage or a potential rebound effect generated by adding an extra furnace, central air conditioning, or additional square footage during the retrofit. Despite the small effects, whole house retrofits still seem to generate more savings than other product upgrades, such as dishwasher, clothes washer, and building envelope (see table 5).

Pool pump upgrades, on the other hand, generate larger savings than the ex ante estimations. This is likely due to the assumptions made during the engineering estimation process. For example, it is assumed that single-family customers upgrade from a two-speed pool pump (a more efficient type) to a variable speed pool pump (the most efficient type) even if they actually upgrade from a single speed pool pump (a less efficient type). It is also assumed that the pump is used for only a few hours a day. These assumptions, approved by CPUC under specific guidelines, lead to underestimation in

28. Examples of previous approved workpapers can be downloaded from CPUC's website, http://deeresources.net/workpapers.

29. An example guideline describing the process utilities have to follow to generate workpapers can be found at http://www.deeresources.com/files/DEER2020/download/CPUC%20WP%20workplan_12242018_Rev1.pdf.

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

Table 5. Comparison of the Impact of Energy Efficiency Upgrade on Electricity Usage with Ex Ante Estimates

	2010–14		
	Changes in Monthly Electricity Usage (kWh)		
Variables	Measured with Billing Data	Utility Engineering Estimates	Realization Rate
Appliance	5.004**	−14.19	. . .
	(2.123)		
Consumer electronics	−28.336	−15.55	. . .
	(20.324)		
HVAC	−14.241***	−21.89	65%
	(1.555)		
Lighting	−.313	−4.25	7%
	(1.001)		
Pool pump	−123.979***	−42.17	294%
	(1.914)		
Refrigeration	−32.335***	−53.25	61%
	(.475)		
Audits	−24.017***	−10.59	227%
	(.657)		
Whole house retrofit	−10.514***	−55.35	19%
	(3.865)		
Building envelope	.770	−5.19	. . .
	(4.944)		
Household-month fixed effects	Yes		
Month-year fixed effects	Yes		
Observations	51,570,259		
R-squared	.9699		

Note. This table reports coefficient estimates and standard errors (in parentheses). The coefficients of interest are indicator variables for households that have participated in the EE upgrade financial incentive programs for that specific product (i.e., lighting, HVAC, etc.). In this table, the dependent variable is the 2010–14 monthly electricity consumption in kilowatt-hours. Estimations are based on matching EE participants with those who have never participated in the programs in the following variables: exact match on building type (single family, multifamily, condo, etc.), vintage bins, climate zone, square footage percentile (only for single and multifamily housing), also fuzzy nearest distance matching on census block group–level variables, such as median income, density, poverty, white, black, Asian, Hispanic, education, age, percentage of ownership, percentage of occupancy rate. Other matching covariates are from the census block group–level variables, such as median income, density, poverty, white, black, Asian, Hispanic, education, age, percentage of ownership, percentage of occupancy rate, and also variables at the account level, including whether homeowner or not, geographic location—x-y coordinates, whether the household is registered under CARE/FERA energy discount programs for low-income households, and whether the household is identified as having a pool or not. Mean pretreatment electricity consumption is 710 kilowatt-hours per month for households that have ever participated in the energy efficiency programs. Standard errors are clustered at the building level. Ex ante savings are reported using the median value. We do not report the “other” category, but to prevent from overlapping program effect, we still control for “other” upgrades (including the water savings kit).

* $p < .1$.

** $p < .05$.

*** $p < .01$.

ex ante predictions. The results indicate that the bias of engineering estimates, which often lack relevant information on individual behavior, can go both ways, validating the importance of conducting EE evaluations with individual meter or billing data.

4. HETEROGENEOUS EFFECTS

4.1. Financial Incentives

We further investigate product upgrade effectiveness based on the level and delivery mechanism of the financial incentives. We report estimates based on the product upgrade effectiveness interacted with the way subsidies were distributed and the level of subsidy (see table A6). For example, some product upgrades offer indirect financial incentives to upstream actors (i.e., product manufacturers) and midstream actors (i.e., retailers or service providers), while others offer downstream incentives that target end users through mail-in or in-store rebates or discount.

Giving away free light bulbs (results in rows 6, 9) generates smaller savings, compared to those that provide partial financial incentives (table A6, rows 6, 9).[31] On average, lighting rebates lead to 3% significant savings, while free lighting upgrades lead to only 0.3%–0.4% savings. The result remains the same when controlling for the quality of light bulbs using the ex ante predicted savings (table A6, col. 2). 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 effects of cost sharing. First, there might be a selection effect (or screening effect) where cost sharing helps select those who need the product more and therefore use more (Ashraf et al. 2010).[32] Second, there might be a psychological effect where 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, there might be a signaling effect where people view the product as having higher quality, thus encouraging its usage (Bagwell and Riordan 1991; Riley 2001). These results about free products only apply to lighting. However, we observe differences regarding the type of subsidy provided for HVAC and pool pump upgrades, which can be provided either as a monetary subsidy for the equipment, or a subsidy for the labor through direct install. The results show that the direct install subsidy is associated with decreases in electricity for pool pumps but not HVAC. Furthermore, the subsidies can be provided to the upstream/midstream (contractor or retailer) or downstream (to the consumer). We find that downstream subsidies are associated with decreases in electricity in the case of lighting and pool pumps.

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

32. 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 the new light bulbs save energy.

Therefore, cost sharing 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 EE upgrades by choosing the appropriate delivery mechanism for different products. The result is not very conclusive comparing upgrades giving incentives to the upstream/midstream entity versus to the end users.[33]

4.2. Household and Building Characteristics

We also conducted subgroup analysis. We estimate equation (2) by income quartile, vintage, square footage, and climate zone to understand where EE upgrades deliver the largest/least savings. This analysis can help policy makers better target upgrades and areas that deliver the largest savings.

We find that the magnitudes of the coefficient for pool pump and refrigeration upgrades are consistent with the main result in table 4, column 4, based on the subgroup analysis. Therefore, we focus our analysis of heterogeneous effects among products that have potential rebound effects—EE participants seem to use more electricity after upgrading, compared to nonparticipants. Then we try to compare electricity usage among different income quartiles, square footage quartiles, and vintage subgroups.

We compare savings 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 (see table A7). We do not see significant differences by income groups across various products, except for audits.[34] Audits 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 minuscule. However, since income is identified under a coarse block group, rather than account or building level, we need to take this result with a grain of salt.

In addition, we compare savings between those who live in a larger home (first and second square footage quartile) and those who live in a smaller home (third and fourth square footage quartile) (see table A7).[35] Most of the comparisons based on the size of

33. Pool pump programs incentivizing end users seem to generate slightly larger savings than incentivizing upstream/midstream manufacturers and contractors. However, the difference may not be considered large in economic terms.

34. Even though the coefficients of HVAC, lighting, 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.

35. For multifamily 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 subgroup analysis.

the building do not yield economically significant differences. The only exception is audits. Audits lead to around 5% savings for large buildings (third and fourth quartile), and 1% savings for smaller homes (first and second quartile). This result may simply be due to the larger savings potential in retrofitting a large building.

We also compare savings between those who live in a home built before 1978 and those in a home built after 1978 (see table A7). We chose 1978 as the cut-off year because this is the year when California Title 24 Building Energy Efficiency Standards were established.[36] The most interesting results here concern HVAC and whole house retrofits. For HVAC, coefficients are positive before 1978 and negative after 1978. This indicates that participants who upgrade their HVAC systems in older buildings use more electricity after upgrading, compared to the nonparticipants. It could be that energy efficient HVAC and building efficiency may complement each other. Indeed, Liang et al. (2017) found that some initial building attributes may affect the effectiveness 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 that those who have participated in both HVAC and whole house retrofits reduced their electricity consumption by 8% more than those who have simply done HVAC retrofit.[37]

Regarding whole house retrofitting, the coefficient is negative before 1978 and positive after 1978. This result indicates that older homes benefit from the retrofitting. This is consistent with engineering assumptions that EE investments in older buildings may yield larger savings potential. However, newer homes increase their consumption after the retrofit. This surprising result could be explained by an increase in the size of the appliances installed or adding an extra unit (such as room air conditioner, central air conditioning, furnace) during the retrofit. Based on anecdotal evidence from one anonymous SCE program manager, some people increase the size of their homes when conducting a whole house retrofit. The government contracted report also found evidence of all the stated potential rebound actions (DNV.GL 2014b).[38]

The results for the whole building retrofits are the opposite. After retrofitting, participants in older buildings consume approximately 2%–3% less electricity, while participants in newer buildings use more electricity. This heterogeneous effect in whole building retrofits indicates that older buildings may have larger savings potential. Nevertheless, we recognize the fact that all whole house building retrofits have also potential to improve

36. 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.

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

38. DNV.GL (2014). Whole House Retrofit Impact Evaluation of Energy Upgrade California Programs Work Order 46, http://www.calmac.org/publications/CPUC_WO46_Final_Report.pdf.

natural gas savings. However, we do not have access to natural gas data for further analysis. This limitation may cause us to underestimate the overall energy savings from whole house retrofits, and the magnitude of this underestimation will depend on the use of natural gas services in homes.

We further examine the impact of different EE product upgrades by climate zone (see table A8). A climate zone is defined based on its weather pattern according to California Energy Commission's definition (see fig. A4). The overall results are consistent with the main results. The energy savings 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 2, which shows the seasonal impact of HVAC and whole house retrofits.[39]

We compare product effectiveness across different building use types (see table A9). The classification is based on building use and construction design. The final categories used in the paper are based on the categorization adopted by Arizona State University Researchers (Reyna and Chester 2015), namely, single family, multifamily, condominium, mixed use (any residential usage mixed with commercial or industrial usage), and residential other (e.g., mobile home). The ranking of the savings is mostly consistent with the overall result—pool pump and refrigeration upgrades are on average the most effective in terms of electricity savings. It seems like building envelope upgrades save much more electricity in other types of residential buildings,[40] compared with single-family, multifamily, and condominium buildings.

4.3. Interaction Effect

Some products may work better when adopted in combination with other technological upgrades. We examine some of the most common combinations to better understand their impact. We find some intriguing results for HVAC and refrigeration upgrades: HVAC upgrades work better when a whole house retrofit is also conducted (see table A10). However, electricity savings are not statistically significant when an HVAC upgrade is done with a building envelope upgrade (though the coefficient is negative). The full interaction results (see table A11) are consistent with the results in table A10. We also find a similar effect for refrigeration, where refrigeration upgrade is more effective with the house retrofitting (see table A10, col. 3). Upgrading refrigeration and retrofitting the whole house together can further decrease 3.9% of electricity consumption, in addition to the original 6% of savings from the refrigeration upgrade (results are consistent in table A11 when we

39. For example, HVAC retrofits deliver 1% savings in climate zone 10, which requires higher demand for energy needed to heat a building (1,678 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 savings results from whole house retrofits also differ from zone to zone.

40. This is driven especially by mixed use buildings.

conduct a similar analysis interacting refrigeration with all other products). Yet, upgrading both refrigeration and building envelope does not show a significant enhancing effect. Finally, lighting upgrade, if done in conjunction with other various upgrades (such as appliance, HVAC, refrigeration, and whole house retrofit) can save more electricity. It is worth noting that HVAC and refrigeration upgrades both work better in combination with whole house retrofit, but the interaction of HVAC and refrigeration upgrades and building envelope upgrade is insignificant. This might be because a whole house retrofit is a more comprehensive home enhancement than a building envelope upgrade. A whole house retrofit aims to configure all aspects of the house to be as energy efficient as possible as well as to make all improvements complement each other. For example, a house retrofit usually ensures that all key issues such as air quality, dampness management, and ventilation are managed appropriately.[41]

As a large amount (44%) of the EE participants upgraded more than one product, these results indicate that the impact of lighting upgrades may be overstated if one fails to control for overlapping upgrade product impact. In summary, the most effective interactions include HVAC, refrigeration, and lighting upgrades conducted with a whole house retrofit.

5. ROBUSTNESS CHECKS

One important assumption for this difference-in-difference estimation is that in the absence of the subsidy, the matched nonparticipants should have a similar electricity consumption pattern to EE upgrade participants after controlling for household-month and time fixed effects–parallel trend assumption. This assumption is not directly testable. Still, we can assess the robustness of our results based on several different tests.

First, we conduct a falsification test, as the recent literature recommends a move away from relying on traditional parallel trend pretests (Bilinski and Hatfield 2019; Freyaldenhoven et al. 2019; Roth 2019; Rambachan and Roth 2021).[42] We randomly assign the installation date two years before the actual reported installation date. As expected, the impact for those effective programs, namely, pool pump and refrigeration upgrade programs, disappears or appears to be very small if we randomly assign two years before the installation date for the EE participants (see table A12).

Second, we conduct another robustness test to control for the group-specific time trend (see table A13) (Mora and Reggio 2012; Bhuller et al. 2013; Dobkin et al. 2018; Rambachan and Roth 2021). Specifically, we apply an alternative "parallel growth" assumption to control for this normalized time trend for each product. In other words, we assume that the acceleration rate of the outcome difference between the control and

41. See https://www.retrofitacademy.org/what-does-whole-house-retrofit-mean-to-me/.

42. The preexisting trends for the most effective products, pool pump and refrigerator, appear to be parallel (see fig. A2).

the treated would have been constant in the absence of the treatment. Using this alternative assumption, we find that the results for the pool pump and refrigeration upgrades are consistent. The savings results for appliance, HVAC, whole house retrofit, and building envelope upgrades are slightly higher than the previous estimation. Nonetheless, they still yield either very small or nonsignificant savings results compared with their ex ante engineering estimates, indicating concerns for a rebound effect.

Third, we conduct a sensitivity analysis based on the approach in Rambachan and Roth (2021), which allows the presence of nonlinear differential trends between treatment and control groups. This method is useful even when the researchers have detected a violation of the parallel trend assumption, as the analysis helps to see how large this nonlinearity is for the detected results to break down. To do this, we estimate an event-study specification under different assumptions about the degree of the nonlinearity—M, where $M = 0$ when the control and treatment groups follow a linear trend. A larger M means a larger change in the differential slope from period to period. Based on our calculation, our savings results are reasonably robust for the most effective upgrades, namely, pool pump and refrigeration, as the breakdown values are both sufficiently large.[43] See some more details in appendix C and figure A5. Finally, we assess the robustness of our results by adding a household-specific time trend variable—a similar test to that conducted by Davis et al. (2014) (see table A14). We explain the details in appendix C.

We also tried different types of matching. We conduct the matching separately by each type of product as some may be concerned that people who adopted different technologies may be different (see table A15). We also conduct two other matching algorithms used by Davis et al. (2014) (see table A16): we match the EE participants with the never participants by only their location through the x-y coordinates (col. 1), and we use location and households' previous electricity usage (col. 2). The results are all quite stable.[44]

We ran 11 separate regressions to estimate savings impact for each type of upgrade one by one without controlling for overlapping upgrades (see table A17). We find that most results are consistent with our main results in table 5, except for lighting. The savings estimations are slightly overestimated if we do not control for other product types. This result again stresses the need to control for overlapping product effect.

Finally, we conduct several other robustness checks. As the actual upgrade may take time to be effective, we allow lags (one month, two month, three month) for the treatment as alternative specification. We also take into account special cases where

---

43. In our case, the breakdown value of M is 0.025 for pool pump, and 0.015 for refrigerator. To put this into economics context, it is equivalent to a 20% (2.5/0.126) change in electricity price, and 12% change in electricity price (1.5/0.126), respectively, if we assume that the short-run price elasticity in electricity demand is 0.126 (Labandera et al. 2017).

44. The slight difference is that in table A15 when products were matched one by one, incentives for whole house retrofitting and building envelope no longer associated with positive coefficients, yet they are not statistically different from zero.

households may have installed solar panels or when households may move to different buildings during our study period. We use not yet participating households as the alternative control group to reestimate the coefficients. All the results are consistent to the main results. See all the details in appendix C and tables A18 to A20.

6. CONCLUSION

Given the large amount of public funds spent on promoting energy efficiency upgrades, our analysis adds credible empirical evidence about the effectiveness of existing EE upgrades. Overall, our results—based on all upgrades providing financial incentives in SCE service territory—point to around 4% energy savings. However, when looked at individually, it is clear that some EE product upgrades are more effective than others. For example, subsidizing pool pump and refrigeration upgrades consistently leads to high rates of energy savings (12% and 6%, respectively), while other upgrades raise concerns over rebound effects. In other words, as evidenced by the literature, customers may use more energy or purchase larger units when the per unit cost of energy service is cheaper (Davis 2008; Houde and Aldy 2017; Sun 2018). Though there is not yet enough information to distinguish the mechanisms behind the rebound effect, we identify building types that exhibit the effect and advise policy makers to keep these in mind when developing incentive programs.

In addition, the results highlight the importance of incorporating end-use billing data in impact evaluations, as we find discrepancies between engineering estimates and actual measured savings. For example, the CPUC's overall impact evaluation report, which is based on engineering estimates, claims that lighting programs could deliver large savings. However, this is not what we find in our analysis. Some of this discrepancy may come from ineffective program implementation rather than problems with the product itself—for example, we find that giving away free light bulbs is associated with nonsignificant savings, while subsidizing the cost of new light bulbs is more effective.

Our study highlights the potential for improving EE upgrade effectiveness by choosing the appropriate subsidy for different products. The results also indicate that policy makers should consider the allocation of program funding not simply based on engineering projections but also based on real measured impacts such as those described in this study. For example, lighting programs, which are seen worldwide as effective methods of increasing energy efficiency, may only have minor effects in regions like California. Measured impacts can be collected by better monitoring new programs and correcting accordingly. Upgrade incentives for HVAC, building envelope, and dishwasher incentives may generate cobenefits such as home comfort and convenience, but progress toward the overall environmental target of reducing electricity consumption may need further investigation.

It is worth noting that one limitation of this analysis results from the categorization of product types, for example, HVAC upgrades do not clearly distinguish between air conditioning units and heating systems. The inability to distinguish two separate energy delivery systems limits our ability to evaluate the effectiveness of each system individually. Therefore, for future evaluation purposes, we recommend creating a more refined classification of the program data. This classification could be informed by the "technology category" as defined in the Building Energy Data Exchange Specification (BEDES), which is a dictionary of terms developed by the US Department of Energy for stakeholders to facilitate important energy investment decisions. We recommend that, within this classification scheme, categories that represent a broad collection of energy services (i.e., heating, ventilation, cooling) 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 United States, natural gas is also widely used. While the analysis of natural gas is beyond the scope of this analysis, it would be helpful to evaluate natural gas in future work. Indeed, it is possible that some households during a remodel decide to switch the energy source of their appliances. However, current market research indicates that this might not be a widespread phenomenon during the time period of our analysis. Nevertheless, our results need to be evaluated with caution, particularly as electricity appliances become more efficient and compete more effectively with gas appliances.

An important complement to our study will be to understand the low uptake of energy efficiency upgrades. In order to implement the most effective programs, policy makers may need to better understand how different program designs affect adoption rates. For example, pool pump upgrades may have great savings potential but will have low adoption potential in a low-income neighborhood without many pools, while HVAC and whole house retrofits may exhibit greater adoption potential for low-income families and generate larger side benefits, such as home comfort. To better understand this distributional effect, future researchers should conduct household-level surveys or experiments to determine, for example, the optimal cost-sharing level for various energy efficiency upgrades.

REFERENCES

Adan, Hassan, and Franz Fuerst. 2016. Do energy efficiency measures really reduce household energy consumption? A difference-in-difference analysis. *Energy Efficiency* 9 (5): 1207–19.

Adekanye, Oluwatobi G., Alex Davis, and Inês L. Azevedo. 2021. Do LED lightbulbs save natural gas? Interpreting simultaneous cross-energy program impacts using electricity and natural gas billing data. *Environmental Research Communications* 3 (1): 015003.

Alberini, Anna, Will Gans, and Charles Towe. 2016. Free riding, upsizing, and energy efficiency incentives in Maryland homes. *Energy Journal* 37 (1): 259–90.

Alberini, Anna, and Charles Towe. 2015. Information v. energy efficiency incentives: Evidence from residential electricity consumption in Maryland. *Energy Economics* 52: S30–S40.

Alix-Garcia, Jennifer M., Katharine R. E. Sims, and Patricia 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.

Allcott, Hunt, and Michael Greenstone. 2012. Is there an energy efficiency gap? *Journal of Economic Perspectives* 26 (1): 3–28.

———. 2017. Measuring the welfare effects of residential energy efficiency programs. Unpublished manuscript.

Allcott, Hunt, and Judd B. Kessler. 2019. The welfare effects of nudges: A case study of energy use social comparisons. *American Economic Journal: Applied Economics* 11 (1): 236–76.

Allcott, Hunt, and Sendhil Mullainathan. 2010. Behavior and energy policy. *Science* 327 (5970): 1204–5.

Arkes, Hal R., and Catherine Blumer. 1985. The psychology of sunk cost. *Organizational Behavior and Human Decision Processes* 35 (1): 124–40.

Asensio, Omar Isaac, and Magali A. Delmas. 2016. The dynamics of behavior change: Evidence from energy conservation. *Journal of Economic Behavior and Organization* 126:196–212.

———. 2017. The effectiveness of US energy efficiency building labels. *Nature Energy* 2 (4): 1–9.

Ashraf, Nava, James Berry, and Jesse M. Shapiro. 2010. Can higher prices stimulate product use? Evidence from a field experiment in Zambia. *American Economic Review* 100 (5): 2383–413.

Aydin, Erdal, Nils Kok, and Dirk Brounen. 2017. Energy efficiency and household behavior: The rebound effect in the residential sector. *RAND Journal of Economics* 48 (3): 749–82.

Bagwell, Kyle, and Michael H. Riordan. 1991. High and declining prices signal product quality. *American Economic Review* 81:224–39.

Bhuller, Manudeep, Tarjei Havnes, Edwin Leuven, and Magne Mogstad. 2013. Broadband internet: An information superhighway to sex crime? *Review of Economic studies* 80 (4): 1237–66.

Bilinski, Alyssa, and Laura A. Hatfield. 2019. Nothing to see here? Non-inferiority approaches to parallel trends and other model assumptions. Unpublished manuscript.

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.

———. 2020. Do energy efficiency investments deliver at the right time? *American Economic Journal: Applied Economics* 12 (1): 115–39.

Chen, Victor L., Magali A. Delmas, William J. Kaiser, and Stephen L. Locke. 2015. What can we learn from high-frequency appliance-level energy metering? Results from a field experiment. *Energy Policy* 77:164–75.

CPUC (California Public Utilities Commission). 2015. 2010–2012 energy efficiency annual progress evaluation report. CPUC, San Francisco.

———. 2016. Regulating energy efficiency—a primer on the CPUC's energy efficiency programs. CPUC, San Francisco.

Davis, Lucas W. 2008. Durable goods and residential demand for energy and water: Evidence from a field trial. *RAND Journal of Economics* 39 (2): 530–46.

Davis, Lucas W., Alan Fuchs, and Paul Gertler. 2014. Cash for coolers: Evaluating a large-scale appliance replacement program in Mexico. *American Economic Journal: Economic Policy* 6 (4): 207–38.

Davis, Lucas W., and Paul J. Gertler. 2015. Contribution of air conditioning adoption to future energy use under global warming. *Proceedings of the National Academy of Sciences* 112 (19): 5962–67.

Davis, Lucas W., Sebastian Martinez, and Bibiana Taboada. 2020. How effective is energy-efficient housing? Evidence from a field trial in Mexico. *Journal of Development Economics* 143:102390.

Dehejia, Rajeev H., and Sadek Wahba. 1999. Causal effects in nonexperimental studies: Reevaluating the evaluation of training programs. *Journal of the American Statistical Association* 94 (448): 1053–62.

———. 2002. Propensity score-matching methods for nonexperimental causal studies. *Review of Economics and Statistics* 84 (1): 151–61.

Delmas, M. A., M. Fischlein, and O. I. Asensio. 2013. Information strategies and energy conservation behavior: A meta-analysis of experimental studies from 1975 to 2012. *Energy Policy* 61:729–39.

DNV.GL. 2014a. HVAC impact evaluation final report WO32 HVAC. Vol. 1, Report. http://calmac.org/publications/FINAL_HVAC_Impact_Evaluation_WO32_Report_28Jan2015_Volume1_Report.pdf.

———. 2014b. Whole house retrofit impact evaluation of energy upgrade California programs Work Order 46. http://www.calmac.org/publications/CPUC_WO46_Final_Report.pdf.

———. 2016. Impact evaluation of 2013–2014 SDG&E residential VSD pool pump program. http://www.calmac.org/publications/Res6_2013-2014_SDGE_VSD_Pool_Pump_Program_Evaluation_FINAL_REPORT_to_CALMAC.pdf.

———. 2017a. Final report: 2015 home upgrade program impact evaluation. http://www.calmac.org/publications/RES_5.1_HUP_FINAL_REPORT_ATR_06-30-17.pdf.

———. 2017b. Universal audit tool impact evaluation: Residential. http://www.calmac.org/publications/EDRes9_UAT_ResReport_CALMAC_final.pdf.

Dobkin, C., A. Finkelstein, R. Kluender, and M. J. Notowidigdo. 2018. The economic consequences of hospital admissions. *American Economic Review* 108 (2): 308–52.

EIA (Energy Information Administration). 2009. Household energy use in California. https://www.eia.gov/consumption/residential/reports/2009/state_briefs/pdf/ca.pdf.

Energy Commission and Public Utilities Commission. 2005. Energy action plan II implementation roadmap for energy policies. http://docs.cpuc.ca.gov/word_pdf/REPORT/51604.pdf.

Fowlie, Meredith, Michael Greenstone, and Catherine Wolfram. 2018. Do energy efficiency investments deliver? Evidence from the weatherization assistance program. *Quarterly Journal of Economics* 133 (3): 1597–644.

Freyaldenhoven, Simon, Christian Hansen, and Jesse M. Shapiro. 2019. Pre-event trends in the panel event-study design. *American Economic Review* 109 (9): 3307–38.

Gillingham, Kenneth, Amelia Keyes, and Karen Palmer. 2018. Advances in evaluating energy efficiency policies and programs. *Annual Review of Resource Economics* 10:511–32.

Giraudet, Louis-Gaëtan, Sébastien Houde, and Joseph Maher. 2018. Moral hazard and the energy efficiency gap: Theory and evidence. *Journal of the Association of Environmental and Resource Economists* 5 (4): 755–90.

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

Graff Zivin, Joshua, and Kevin Novan. 2016. Upgrading efficiency and behavior: Electricity savings from residential weatherization programs. *Energy Journal* 37 (4): 1–23.

Guardian. 2009. America's new green guru sparks anger over climate change U-turns, May 23. https://www.theguardian.com/environment/2009/may/24/steven-chu-environmentalist-anger.

Ho, Daniel E., Kosuke Imai, Gary King, and Elizabeth A. Stuart. 2007. Matching as nonparametric preprocessing for reducing model dependence in parametric causal inference. *Political Analysis* 15 (3): 199–236.

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): 1171–81.

Houde, Sébastien, and Joseph E. Aldy. 2017. Consumers' response to state energy efficient appliance rebate programs. *American Economic Journal: Economic Policy* 9 (4): 227–55.

Imbens, Guido W., and Jeffrey M. Wooldridge. 2009. Recent developments in the econometrics of program evaluation. *Journal of Economic Literature* 47 (1): 5–86.

Labandeira, Xavier, José M. Labeaga, and Xiral López-Otero. 2017. A meta-analysis on the price elasticity of energy demand. *Energy Policy* 102:549–68.

Liang, Jing, Yueming Qiu, Timothy James, Benjamin L. Ruddell, Michael Dalrymple, Stevan Earl, and Alex Castelazo. 2017. Do energy retrofits work? Evidence from commercial and residential buildings in Phoenix. *Journal of Environmental Economics and Management* 92:726–43.

Liddle, Brantley, Tian Sheng Allan Loi, Anthony D. Owen, and Jacqueline Tao. 2020. Evaluating consumption and cost savings from new air-conditioner purchases: The case of Singapore. *Energy Policy* 145:111722.

Mathew, Paul A., Laurel N. Dunn, Michael D. Sohn, Andrea Mercado, Claudine Custudio, and Travis Walter. 2015. Big-data for building energy performance: Lessons from assembling a very large national database of building energy use. *Applied Energy* 140:85–93.

McKinsey and Company 2009. Unlocking energy efficiency in the U.S. economy. Technical report, McKinsey Global Energy and Materials.

Mora, Ricardo, and Iliana Reggio. 2012. Treatment effect identification using alternative parallel assumptions. Unpublished manuscript.

Novan, Kevin, and Aaron Smith. 2018. The incentive to overinvest in energy efficiency: Evidence from hourly smart-meter data. *Journal of the Association of Environmental and Resource Economists* 5 (3): 577–605.

Pincetl, Stephanie, Robert Graham, Sinnott Murphy, and Deepak Sivaraman. 2016. Analysis of high-resolution utility data for understanding energy use in urban systems: The case of Los Angeles, California. *Journal of Industrial Ecology* 20 (1): 166–78.

Pincetl, Stephanie, and LA Energy Atlas Development Team. 2015. LA energy atlas. California Center for Sustainable Communities, University of California, Los Angeles.

Qiu, Yueming, and Anand Patwardhan. 2018. Big data and residential energy efficiency evaluation. *Current Sustainable/Renewable Energy Reports* 5 (1): 67–75.

Rambachan, Ashesh, and Jonathan Roth. 2021. An honest approach to parallel trends. Unpublished manuscript.

Reyna, Janet L., and Mikhail V. Chester. 2015. The growth of urban building stock: Unintended lock-in and embedded environmental effects. *Journal of Industrial Ecology* 19 (4): 524–37.

Riley, John G. 2001. Silver signals: Twenty-five years of screening and signaling. *Journal of Economic literature* 39 (2): 432–78.

Rosenbaum, Paul R., and Donald B. Rubin. 1985. Constructing a control group using multivariate matched sampling methods that incorporate the propensity score. *American Statistician* 39 (1): 33–38.

Roth, Jonathan. 2019. Pre-test with caution: Event-study estimates after testing for parallel trends. Unpublished manuscript.

Rubin, Donald 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.

———. 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.

Scheer, Jim, Matthew Clancy, and Sadhbh Ní 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, Elizabeth A. 2010. Matching methods for causal inference: A review and a look forward. *Statistical Science* 25 (1): 1–21.

Sun, Bixuan. 2018. Heterogeneous direct rebound effect: Theory and evidence from the Energy Star program. *Energy Economics* 69:335–49.

Thaler, Richard. 1980. Toward a positive theory of consumer choice. *Journal of Economic Behavior and Organization* 1 (1): 39–60.

Trotta, Gianluca. 2018. The determinants of energy efficient retrofit investments in the English residential sector. *Energy Policy* 120:175–82.