

APPLICATION PAPER 

Limitations of econometric evaluation of nonrandomized residential energy efficiency programs: A case study of Northern California rebate programs

Evan D. Sherwin^{1,*} , Russell M. Meyer² and Inês M.L. Azevedo^{1,3,4} 

¹Department of Energy Resources Engineering, Stanford University, Stanford, California, USA

²Oracle Corporation, Santa Monica, California, USA

³Woods Institute for the Environment, Stanford University, Stanford, California, USA

⁴Precourt Institute for Energy, Stanford University, Stanford, California, USA

*Corresponding author: E-mail: evands@stanford.edu

Received: 14 June 2021; **Revised:** 08 October 2021; **Accepted:** 08 November 2021

Keywords: Econometrics; energy efficiency; ex-post evaluation; opt-in; selection bias

Abstract

Residential energy efficiency programs play an important role in combating climate change. More precise quantification of the magnitude and timing of energy savings would bring large system benefits, allowing closer integration of energy efficiency into resource adequacy planning and balancing variable renewable electricity. However, it is often difficult to quantify the efficacy of an energy efficiency intervention, because doing so requires consideration of a hypothetical counterfactual case in which there was no intervention, and randomized control trials are often implausible. Although quasi-experimental econometric evaluation sometimes works well, we find that for a set of energy efficiency rebate programs in Northern California, a naïve interpretation of econometric measurement finds that rebate participation is associated with an average *increase* in electricity consumption of 7.2% [4.5%, 10.1%], varying in magnitude and sign depending on the type of appliance or service covered by the rebate. A subsequent household survey on appliance purchasing behavior and analysis of utility customer outreach data suggest that this regression approach is likely measuring the gross impact of buying a new appliance but fails to adequately capture a counterfactual comparison. Indeed, it is unclear whether it is even possible to construct a suitable counterfactual for econometric analyses of these rebate programs using data generally available to electric utilities. We view these results as an illustration of a limitation of econometric methods of program evaluation and the importance of weighing engineering modeling and other imperfect methods against one another when attempting to provide useful evaluations of real-world policy interventions.

Impact Statement

Many energy utilities offer residential energy efficiency rebate programs to reduce energy consumption and resulting environmental impacts. We find that for a particular set of rebate programs for energy efficient household appliances and services, common econometric methods find that participating households tend to increase electricity consumption after applying for rebates. Thus, it might appear that these efficiency programs did not actually save energy. However, additional utility data and a household survey suggest that the observed increase was likely measuring the “effect” of buying a new appliance. In such circumstances, energy savings

 This research article was awarded an Open Materials badge for transparent practices. See the Data Availability Statement for details.

© The Author(s), 2022. Published by Cambridge University Press. This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted re-use, distribution, and reproduction in any medium, provided the original work is properly cited.

estimates based on engineering models may be more appropriate than econometric methods. This illustrates the importance in policy evaluation of picking the right quantitative tool for the job.

1. Introduction

Residential energy efficiency is a major component of national- and state-level energy policies in the United States (Sweeney, 2016). Since 2005, the U.S. federal government has spent over \$13 billion on residential energy efficiency programs (Borenstein and Davis, 2016), whereas state-level utility spending was \$8.4 billion in 2019 alone (Berg et al., 2020). Energy efficiency strategies in the residential sector are often found to be the most cost-effective climate mitigation strategies, with numerous studies and analyses that estimate both the potential and achieved cost-effective savings from residential energy efficiency programs (Meier et al., 1982; Koomey et al., 1991; Rosenfeld et al., 1991, 1993; Rubin et al., 1992; Blumstein and Stoff, 1995; Jackson, 1995; Levine et al., 1997; Brown et al., 1998; Rosenfeld, 1999; Coito and Rufo, 2002; Nadel et al., 2004; McKinsey, 2007, 2009; Goldstein, 2008; Richter et al., 2008; Ürge-Vorsatz et al., 2009; NRC, 2010; Azevedo et al., 2013). A review by Saunders et al. (2021) highlights the findings from research in energy efficiency in the last 40 years and stresses that key uncertainties persist regarding the outcomes of energy strategies and programs (Saunders et al., 2021).

With increasing levels of renewable energy deployment in electric power systems across the world and the inherent supply variability of those resources, the timing of electricity demand and the timing of savings are increasingly important (Boomhower and Davis, 2020). For example, abundant midday solar electricity in California creates substantial periods of negative wholesale pricing in the spring and the fall, resulting in peak demand periods that are pushed back until 8 or 9 pm at night, when little to no solar is available (Bajwa and Cavicchi, 2017; PG&E, 2019). As a result, a kilowatt hour (kWh) of electricity saved in the evening provides many more system benefits, including emissions reductions, than the same kWh saved on a spring afternoon. Of course, the extent to which electric power transmission, distribution, and generation capacity planners can incorporate these time-based savings into their planning decisions depends on how well we can measure them.

Energy savings from energy efficiency programs cannot be measured directly—there is no way to directly measure something that did not happen—so energy efficiency program evaluators must rely on engineering or econometric methods to estimate energy savings. The majority of residential energy efficiency activity in the United States has been designed, by necessity, as opt-in programs. A major exception is home energy reports, which many utilities send by default to customers to provide feedback on their energy consumption with the aim of encouraging behavioral change (Allcott and Kessler, 2019). By their nature, opt-in programs require a homeowner, landlord, building manager, or occupant to make an active decision to participate. As a consequence, all opt-in programs have some degree of unavoidable selection bias: The group of participants who elect to engage in an energy efficiency program will be different from the group that does not elect to engage. We would reasonably expect that the two groups—those that would elect to participate in an energy efficiency program and those who would not—will have different future energy consumption patterns even in the absence of an energy efficiency program intervention. Participants in opt-in programs are potentially more likely, as a group, to engage in energy-saving behavior in the absence of a utility-sponsored efficiency program. They may also be more receptive to other messaging about the environmental or financial benefits of saving energy, more cognizant of their own energy consumption patterns, or simply have fewer hurdles to engaging in energy-saving actions.

The extent to which any of the (usually unobservable) differences between opt-in participants and nonparticipants are correlated with future energy consumption patterns is a challenge for attribution: How can program managers and regulators estimate the marginal effect (the “additionality”) that energy

efficiency programs are creating? What would the group that opted into the program have done in the absence of the program, and how great were energy savings induced by program expenditures?

These estimates of savings are important, because they are crucial for evaluating the performance and cost-effectiveness of programs, and for comparison against other potential uses of scarce societal resources. For utilities and program implementers, they are generally used to measure progress toward regulatory energy efficiency requirements. Significant effort has been devoted to creating measurement standards that can be used for estimating opt-in efficiency program progress. The International Performance Measurement and Verification Protocol (IPMVP) aims to provide a “flexible framework of measurement and verification options” that “adhere to the principles of accuracy, completeness, conservativeness, consistency, relevance, and transparency” (EVO, 2012). The Uniform Methods Project (UMP) is a U.S. Department of Energy effort based on the IPMVP, but which is scoped to provide “a more detailed approach to implementing” the options from that protocol (Li et al., 2017).

At a high level, there are two broad categories of estimation methods for opt-in efficiency programs. *Engineering* estimates (IPMVP options A, B, and D) simulate the effect of using a more efficient appliance or adding building improvements, such as insulation, compared with a less efficient counterfactual case (ACEEE, 2019). Utilities and regulatory agencies in more than 25 U.S. states publish Technical Reference Manuals (TRMs) based on such estimates, to generate “deemed” values of clearly defined efficiency activities that are applied toward regulatory energy efficiency mandates (see Li and Dietsch, 2017 for further details). However, engineering estimates, such as those in the TRMs used in many U.S. states, are necessarily somewhat coarse and usually ignore considerable uncertainty in the parameters that affect savings values (Meyer, 2014). They can provide useful insight into the average expected savings from an intervention, perhaps accounting for the regional climate, the type of new appliance, and some characteristics of the residence (NYSJU, 2019). Engineering models also cannot easily include behavioral effects and other potentially critical particularities in a given intervention.

Econometric estimates have long used household electricity consumption data and quasi-experimental approaches to estimate the effects of individual energy efficiency programs. As early as 1986, Fels introduced a weather-normalized regression-based baseline method of energy efficiency evaluation, comparing electricity consumption from monthly utility billing data before and after an intervention, in some cases comparing treatment and control groups (Fels, 1986). Such methods have the potential to capture behavioral effects and other operational characteristics that engineering models cannot generally consider, but only if there is a valid counterfactual. There are now many such econometric methods (including those detailed by Berger and Ucar, 2013) with the U.S. Department of Energy’s UMP outlining standards for such energy efficiency evaluation techniques (Li et al., 2017). Several studies (including Allcott and Greenstone, 2017; Fowlie et al., 2018) suggest that realized energy efficiency savings may be substantially lower than econometric modeling-based estimates.

For econometric studies, the most robust counterfactuals are generated by a randomized controlled trial (RCT) design. Such a method measures the net effect of a program, capturing behavioral as well as engineering components, although they cannot be easily disentangled with this approach. Unfortunately, most residential efficiency programs require active enrollment by participants, as discussed above, which means that an RCT is not possible for those programs.

When RCT design is infeasible, a common evaluation alternative is to employ a quasi-experimental approach. For example, Fowlie et al. (2018) employ a randomized encouragement design (RED) to construct an instrument of the effect of encouragement for the program. Another alternative is to use propensity score matching (PSM) to construct a synthetic control group for comparison to the treatment population in the post-treatment period, as in Qiu and Kahn (2018).

When a RED program is measured on an intention-to-treat basis, it mimics an RCT in its measurement (presuming that the option for participation remains open for the unencouraged control group). This is rarely done, however. More commonly, program administrators and evaluators are interested in estimating a local average treatment effect, which is a measure of the effect of the program on the program participants. This relies on the assumption that participation in the program, subsequent to

program encouragement, “is orthogonal to any factors that impact [energy] consumption” (Hahn and Metcalfe, 2021). This is a difficult hurdle to conclusively overcome, since receptiveness to program encouragement could plausibly be associated with other unobserved characteristics that are associated with future energy consumption. Similarly, unobserved (and, often, unobservable) characteristics present a challenge to estimates generated by PSM. For a synthetic control group created by PSM, balancing all observable characteristics does not, and cannot, guarantee that there are no remaining unobserved factors associated with energy use that are not statistically imbalanced between the treatment and synthetic control groups.

In some cases, quasi-experimental methods, which attempt to construct a counterfactual based on historically occurring, plausibly quasi-random variation in the data, can produce reasonable estimates of causal effects. In the case of energy efficiency program evaluation, Boomhower and Davis (2020) estimated energy savings from a central air conditioner replacement program using hourly electricity consumption data from advanced metering infrastructure (AMI) in Southern California; however, they caution that these results are not necessarily causal due to the quasi-experimental estimation approach, stemming from the program’s design. Novan and Smith (2018) apply a similar analysis in the Sacramento area. Because participating households in both studies had central air conditioning before participating in the program, average changes in pre- and post-replacement electricity consumption (with a regression accounting for appropriate control variables) give a plausible estimate of the net effect of the program. Furthermore, hourly data allowed examination of differential effects at different hours. These hourly estimates found nighttime electricity savings that far exceeded engineering estimates, providing insight into household behavior, namely a preference among households in Southern California to run air conditioners at night (Boomhower and Davis, 2020).

Empirical estimates of energy savings by time of day and season raise the prospect of transforming energy efficiency into a resource that can reliably contribute to resource adequacy planning, integration of variable renewable energy, and possibly even electric power capacity markets. These are among the stated goals of emerging data-driven energy efficiency measurement and verification companies such as OpenEEMeter and the related CalTRACK program (Recurve, 2020).

We apply regression analysis to a class of energy efficiency programs: rebates for efficient appliances and other residential energy efficiency measures in Northern California. Our initial evaluation of the hypothesis that participation in an energy efficiency program is associated with a subsequent decrease in electricity consumption yielded counterintuitive results. Based on these results, we conducted a household survey to assess possible explanations for these results due to appliance purchasing and disposal and conducted detailed discussions with utility employees familiar with the inner workings of these rebates. We find that constructing a defensible counterfactual is difficult, if not impossible, for most opt-in energy efficiency rebate programs. Some of these concerns would be mitigated if the datasets would include other detailed aspects of participant and nonparticipant behavior, such as purchases and retirements of appliances and equipment, or other behavior changes. In many cases, moving forward with a conventional quasi-experimental econometric specification results in estimates of an increase in electricity consumption, rather than savings. We view these results as an illustration of a limitation of econometric methods of program evaluation and the importance of weighing engineering modeling and other imperfect methods against one another when attempting to provide the most useful possible evaluation of a real-world policy intervention.

2. Data and Methods

Our dataset is one of the earliest AMI large datasets, provided by Pacific Gas and Electric Company (PG&E) via the Wharton Customer Analytics Initiative. We applied quasi-experimental energy efficiency evaluation techniques to this dataset. The data include a mix of 15-min and hourly electricity consumption readings, which we aggregate to hourly and ultimately to daily resolution for consistency and computational tractability, for up to 4 years for associated households (Sherwin and Azevedo, 2020). The data

represent a regionally stratified random sample of roughly 30,000 PG&E customer accounts, together with dates for rebate application by type of appliance or service, rebate approval, and check disbursement information for energy efficiency rebates for numerous appliances, services, and building improvements, as well as other important contextual information, such as enrollment in other utility programs. In Section A1 in the Supplementary Material, we provide further details about the dataset.

We use electricity consumption as the dependent variable, with detailed treatment information, pre- and post-treatment data, and dwelling-level and time fixed effects, in a traditional difference-in-difference model of the sort employed both in energy efficiency evaluation and in many fields of applied economics (Qiu and Kahn, 2018; Burlig and Wolfram, 2020).

While we do not directly observe household address information (for data privacy protection), PG&E linked household pseudoaccount identifiers with U.S. Census block information in the provided dataset. Using this location information, we include local hourly temperature. We also observe enrollment information for several other utility programs offered during the study period (see Section A1.6 in the Supplementary Material for data on enrollment in other programs). The data do not include household demographic information, which we supplement with data at the neighborhood-average census block level. See Table A1 and Sections A1.1 and A1.6 in the Supplementary Material for demographic statistics as well as details on enrollment in other utility programs and tariff structures, such as the California Alternate Rates for Energy low-income subsidy.

3. Interval Electricity Consumption Data

Our primary data source is interval electricity consumption data from dwellings associated with approximately 30,000 PG&E residential customer accounts, roughly 10,000 from each of the three regions within the sample, the Central Valley, Inland Hills, and Coast. See Figure A1 and Section A1.1 in the Supplementary Material for further details. In all, 30,349 dwellings had valid electricity consumption readings, meaning that some accounts were associated with multiple dwellings because the household either moved or owned multiple dwellings simultaneously. Although the data were originally provided at 15-min resolution, we aggregated to hourly resolution to merge with temperature data and then, due primarily to computational constraints, aggregated to daily resolution using a degree day-like metric described in Equations (1) and (2).

Interval data collection began only after the deployment of AMI, which was staged beginning largely in the Central Valley in 2007, moving to the Inland Hills, concluding on the Coast. See Figure A2 and Section A1.1 in the Supplementary Material for further details. As a result of this staging, the panel is unbalanced. However, we do not believe that this substantially influenced our results, which are similar in all three regions. See the discussion surrounding Table A4 and Section A2 in the Supplementary Material for further details.

We also use census block location information to approximate local temperature at each dwelling as the weighted average of the hourly temperature at the three weather stations closest to the center of that dwelling's census block, using data from the National Oceanographic and Atmospheric Administration (Menne et al., 2012). We approximate heating and cooling demand using Equations (1) and (2), based on the deviation of the daily high and low hourly temperature, $T_{h,i,t}$ and $T_{l,i,t}$, from 18°C (~65°F), a common set point for analysis of heating and cooling in the United States, setting the deviation to zero if the high temperature is below 18°C or the low temperature is above 18°C (EPA, 2016). This is a rough approximation of degree days, which are common in monthly billing analysis, or of a similar piecewise linear representation of hourly temperature, which becomes possible with hourly data.

$$Temp_{h,i,t} = \max[T_{h,i,t} - 18^{\circ}\text{C}, 0]. \quad (1)$$

$$Temp_{l,i,t} = \max[18^{\circ}\text{C} - T_{l,i,t}, 0]. \quad (2)$$

We conducted our analysis at the dwelling level. We were able to control for all utility programs a household was enrolled in using account-level data, which apply to all dwellings associated with an account. See Section A1.6 in the Supplementary Material for further description of other utility programs available to households during the study period. Rebate participation, including application date, approval date, and check issuance date, were reported at the dwelling level.

4. Difference-in-Difference Regression

We use a difference-in-difference regression approach to measure the association between energy efficiency rebate participation and electricity consumption using Equation (3).

$$\ln(kWh_{i,t}) = \alpha + \beta_j(Temp_{i,t})_j + \gamma(Rebate_{i,t}) + \delta_k(Time_t)_k + \zeta(TimeTrend_t) + \varphi_q(Program_{i,t})_q + \psi_q(Rebate_{i,t}*Program_{i,t})_q + u_i + \varepsilon_{i,t}. \quad (3)$$

The main analysis uses Equation (3), which controls for enrollment in other utility programs and potential interactions between rebate participation and enrollment in these programs. $\ln(kWh_{i,t})$ is the natural logarithm of electricity consumption in kWh, for dwelling i in day t . We use this approach because the distribution of electricity consumption is approximately lognormal and results are interpretable in percentage terms. See Section A5 in the Supplementary Material for further details. The primary coefficient estimate of interest is associated with $Rebate_{i,t}$, which is an indicator variable for dwellings following their first rebate application. We assume that any change in energy consumption associated with efficiency measures begins at roughly the same time as rebate application. We believe this is reasonable, because the current deadline for rebate submission is 60 days after purchase, and households that apply for rebates have already purchased the relevant appliances or efficiency services (PG&E, 2017). $(Temp_{i,t})_j$ is a set of linear and quadratic temperature controls, j , where j has four values, representing daily high and low temperatures for each household, based on Census block location, in a linear and quadratic form. Both high and low temperatures are derived from an average of the three nearest weather stations, represented as the absolute value of the deviation from 18°C, truncated at zero below for high temperatures and above for low temperatures. See Equations (1) and (2) for further details. $(Time_t)_k$ is a set of k indicators for periodic time intervals (months of the year, and days of the week). $TimeTrend_t$ is a linear time trend that is fitted to the model to capture secular changes in electricity consumption over the period of observation, unrelated to the variable of interest. $(Program_{i,t})_q$ represents the q additional PG&E programs, described in Section A1.6 in the Supplementary Material. The model also includes a set of q interaction terms between rebate program participation and the other PG&E programs. The terms α and u_i are the intercept and the dwelling-specific fixed effect. $\varepsilon_{i,t}$ is an unobserved error term.

Figure 1 uses Equation (4), which differentiates between the different types of rebates available.

$$\ln(kWh_{i,t}) = \alpha + \beta_j(Temp_{i,t})_j + \gamma_l(Rebate_{i,l,t}) + \delta_k(Time_t)_k + \zeta(TimeTrend_t) + u_i + \varepsilon_{i,t}. \quad (4)$$

All regression components are identical except that controls related to other utility programs are excluded and rebates are differentiated by type, l . In the “All rebates” case, l corresponds to all rebates. Otherwise, l corresponds to each distinct type of rebate.

See Section A2 in the Supplementary Material for robustness checks, including alternative subsamples and regression specifications.

The household survey was conducted using a separate population of California households, recruited using Amazon Mechanical Turk. The purpose of this survey, which was not linked to electricity consumption data, was to gain insight into household appliance purchasing and disposal behavior and the use of rebates. Such results may not fully generalize to the population in the main analysis, because there may be demographic or other differences between the surveyed population and the sampled population within the PG&E service territory. For more information about the household survey, see Sections A4 and A6 in the Supplementary Material.

5. First Econometric Impressions and Why They Are Misleading

Applying the difference-in-difference regression described above in Equation (3), we find that rebate participation did not appear to reduce electricity consumption and was instead associated with an average increase in electricity consumption of 7.2% with a 95% confidence interval of [4.5%, 10.1%]. Using the simpler regression specification in Equation (4), which does not control for enrollment in other programs, this falls slightly to 6.1% [3.4%, 8.8%]. Our initial hypothesis was that household energy consumption would decrease following efficiency rebate participation. See Section A2 in the Supplementary Material for full regression results.

This increase in electricity consumption appears to be largely attributable to rebates for new appliances, 45% of all rebates, which showed an even higher increase of 9.7% [5.9%, 13.7%], as shown in Figure 1, based on Equation (4) in the Difference-in-Difference Regression section. Differentiating by rebate type using Equation (4), there was a nonsignificant decrease of $-6.1%$ [$-13.7%$, $2.7%$] in electricity consumption for appliance rebates that required recycling of an old appliance. Building shell and unknown/unclassified rebates also show significant increases in electricity consumption of 14.0% [0.7%, 29.0%] and 3.6% [0.2%, 7.1%]. In no case did we see a significant reduction in electricity consumption associated with rebate participation. These results held for a wide array of robustness checks, described in Section A2 in the Supplementary Material. See Section A3 in the Supplementary Material for a detailed breakdown of rebate applications by type over time.

These results could be interpreted naïvely as suggestive evidence that rebates were acting as a subsidy, encouraging households to purchase new, efficient appliances while keeping older, less efficient versions running. The fact that there was no increase in consumption for appliance rebates that required recycling could be construed as evidence supporting this hypothesis. Of course, one must acknowledge a number of potential sources of selection bias correlated with both program participation and energy consumption,

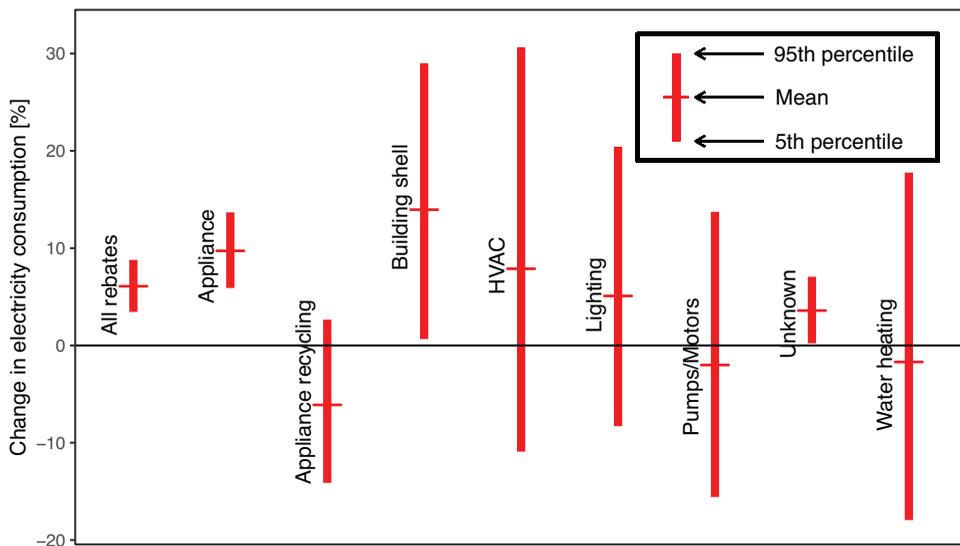


Figure 1. Estimated association between rebate participation and subsequent changes in electricity consumption. Note that rebate participation was associated with a significant increase in electricity consumption that does not appear for appliances that required recycling of an old, less efficient appliance. These results use Equation (4). A naïve interpretation could interpret this as suggestive evidence that energy efficiency rebates lead to increased consumption. However, this result is most likely due to the fact that it is essentially impossible to develop a statistically valid counterfactual for the type of rebate program evaluated in this study, particularly using data generally available to electric utilities.

including the possibility of simultaneous and unmeasured changes in household size, income or employment status, or the household appliance stock or building envelope. However, it is not uncommon for studies with similar statistical limitations and apparently unintuitive results to be published with the aim of at least sparking important discussion, perhaps motivating further, more detailed studies in the future.

The dataset used above does not include important behavioral data. Perhaps households tended to get new efficient appliances simultaneously with changes in household size or major renovations, which would also affect electricity consumption. To what extent did the rebate influence whether a household decided to buy a new appliance, or to buy a more efficient model than they would have otherwise? We procured more data to assess the extent to which the observed increase in electricity consumption was due to households buying a new appliance and keeping an old version.

PG&E graciously shared additional details about the rebates and other efficiency measures employed by households in our sample, described in Section A1.4 in the Supplementary Material. These data clarified that the vast majority of appliance rebates, roughly 75%, were clothes washers, with roughly 15% dishwashers, both of which are appliances that a household is likely to have either zero or one of. Of appliance recycling rebates, over 90% were for refrigerators or freezers. There were some conflicts between the classifications in the original and additional data, with some rebates labeled as “Appliance recycling” in the original dataset apparently not indicating recycling in the additional data. Such data consistency issues are common in many forms of data generated for administrative purposes.

Positive and significant coefficients for building shell efficiency rebates, associated with an average increase in consumption of 14.0% [0.7%, 29.0%], also motivate similar hypotheses. Households installing building shell efficiency measures may be simultaneously expanding other parts of the building or otherwise taking action that may increase overall energy consumption. However, these building shell retrofits constitute only 99 of 5,484 total efficiency rebates in the database, compared with 2,429 appliance rebates without recycling and 470 with recycling requirements. As a result, the remainder of this study focuses on appliance rebates.

6. Household Survey Debunks “Keeping Old Appliance” Hypothesis

The data from PG&E did not include information necessary to understand appliance purchasing behavior. We conducted an online survey of 665 California households, not linked to the provided household-level electricity consumption data, to gain insight into such behavioral factors. The survey is described in detail in Section A4 in the Supplementary Material. We asked these respondents what appliances they had purchased over the past 10 years, whether they had applied for rebates, and whether they already had old versions of the same appliances and if so, what they did with them after buying new ones. We also asked whether and when they had made major renovations to their home, experienced a change in household size, or enrolled in the California Alternate Rates for Energy low-income subsidy, which could increase consumption by reducing the effective price of electricity.

The household survey was motivated by the hypothesis that households that participated in appliance rebate programs (a) kept an old, less efficient version of the same appliance, or (b) purchased appliances they did not already possess. After the first round of data collection, 101 respondents, we added questions to assess the hypotheses that households purchase appliances at the same time as (c) increases in household size or (d) major building renovations or additions. After the conclusion of the survey, we generated hypotheses that (e) households use rebates to purchase more consumptive, but more efficient appliances, for example, refrigerators with ice-makers or (f) households purchase additional appliances or equipment at the same time as any rebates. We were not able to assess hypotheses (e) and (f) in this study.

We found that only 8% of the 222 households that reported getting a rebate for an efficient appliance also reported keeping an old model. Sixty-two percent of applying households had an old and functioning version of the same appliance, but may not have kept it after purchasing the new appliance. Thus, that 38% of rebate applicants did not have an old version. Of those who applied for a rebate for an appliance they previously had in the home, only 12% reported keeping the old version, with the rest recycling (48%),

scrapping (15%), or selling (25%) the old version. Thus, one of our early hypotheses—that households were keeping old, inefficient appliances after getting new, efficient ones—was not supported by our new data.

The survey results suggested that simultaneous home renovations or changes in household size were not the cause of the observed increase in electricity consumption. These questions were added to the survey after we had collected the first 101 responses. Only 13 of the responding 564 households that were asked questions about changes in household size report applying for a rebate within the same period as an increase in household size, whereas 13 report a decrease. This suggests that increases in household size do not explain our regression results. Only 23 of the same 564 households report simultaneous renovations. Many of these renovations include efficiency upgrades such as greater insulation or more efficient windows. As a result, it is likely that this effect is small and its direction is ambiguous. Note that although the survey includes questions about building renovations and associated energy efficiency measures, it does not include questions that would allow us to evaluate hypotheses surrounding the observed increase in electricity consumption following building shell renovations, because the survey focused on appliance purchasing and disposal behavior. See Section A4 in the Supplementary Material for further information.

The most plausible remaining explanation was that households were using rebates to purchase appliances they did not already possess, particularly clothes washers. It was also possible that households were purchasing more efficient versions of appliances with more features than their old versions, or that they were purchasing other new appliances at the same time as the efficient appliances. Unfortunately, the survey did not include questions that would have allowed us to assess these hypotheses.

7. Rebates Only Advertised at Point of Sale

Further examining the question of how households were informed about rebates, we assessed the extent to which rebates could play a role in household purchasing decisions. Analysis of customer communications in the PG&E dataset did not include any evidence of proactive outreach by phone, email, or physical mail about the various rebates available. This means that rebates were primarily advertised at the point of sale, thus reducing the likelihood that customers would even be aware of the existence of rebates until they were at a store selecting new appliances, present at a store for another reason and visiting the appliances section, or in contact with a contractor or repair company.

Thus, many households that took advantage of rebates for efficient appliances or services had likely decided to make a purchase before the rebate could affect their decisions. This means that rebates probably did not spur households to purchase appliances they would not have bought otherwise, an assumption implicit in our initial interpretation of our results. Rebates may have then encouraged households to opt for a more efficient option, but either way, this poses a major selection bias concern for which it is difficult to correct.

8. What Is the Counterfactual?

If the treatment group is households that purchase a new efficient appliance or efficiency service and apply for a rebate, what is the appropriate comparison against which to measure their energy savings? The method we had employed thus far essentially set the control group as “all households that did not apply for a rebate,” including pre-rebate data for households that did. One could easily imagine that households buying a new appliance that they did not previously own would tend to have a subsequent increase in electricity consumption. Thus, the increase in electricity consumption observed in our original regressions could simply be the effect of purchasing a new appliance not previously present in the dwelling, one of the hypotheses that our survey was not able to fully address.

The mental model implicit in interpreting the econometric results in this way is that in the absence of rebates, households would have continued to use the same appliances and household energy services as before. However, if rebates are primarily affecting purchasing decisions at the point of sale, many of the households that applied for rebates in our sample could have been shopping for a new appliance before

learning about the rebates. Thus, it is likely they would have bought a new appliance with or without the rebate. To the extent that these purchasers would have purchased a qualifying efficient appliance anyway, they can be considered free riders to the rebate program. However, even if these participants would have purchased a competing, less efficient model, the appropriate comparison for their post-purchase energy consumption patterns is not their pre-purchase energy consumption patterns. Instead, the ideal comparison is their hypothetical (and unobservable) post-purchase consumption patterns in the absence of the rebate (and the effect that the rebate had on their purchase decision-making). If rebates indeed induce purchase of less consumptive appliances, such a comparison would likely show a decline, not an increase, in energy consumption for at least a substantial fraction of the participants in this sample.

Perhaps a more appropriate control group would be households that purchased a new appliance, or considered purchasing a new appliance, but did not apply for a rebate. However, even assuming one could assemble such data, and doing so would be a substantial endeavor in itself, there could be numerous reasons why a household opted not to take advantage of available energy efficiency rebates, advertised at the point of sale. Such households could have lower incomes, rendering the additional capital expense of more efficient appliances prohibitive even with a rebate. Such households could also be less concerned about energy consumption or could place a high premium on specific features that do not happen to be available in rebate-eligible models. These and many other potential confounding factors could substantially bias the results in ways that are difficult to predict.

In addition, household electricity consumption data tell us nothing about what appliance the household would have purchased in the absence of a rebate. We do not know what appliance the household would have purchased otherwise. Asked directly, the residents themselves could likely only give a general idea of whether and how much the availability of rebates affected their purchasing decisions or would affect future decisions. A recent analysis of U.S. appliance purchasing trends suggests that the effect is relatively small, finding that 70% of participants in the 2009 expansion of U.S. energy efficiency rebate programs were inframarginal, and households would have bought the same appliance without the rebate (Houde and Aldy, 2017). In our view, it would be extraordinarily difficult to match such appliance sales figures to household electricity consumption, control for myriad confounding factors, and produce an estimate of the resulting energy savings that is more credible than existing engineering estimates.

However, it is unclear how even a randomized experiment could satisfactorily address the fundamental question of how much energy is saved through appliance rebate programs relative to what consumption would be in the absence of the programs. One way to conduct such an experiment would be through a RED, in which a randomly selected subset of households is given promotional materials informing them of the existence of rebates, perhaps even limiting rebate availability to these selected households (e.g., Fowlie et al., 2018; Hahn and Metcalfe, 2021).

In such an experiment, the question of the counterfactual remains. One could get an unbiased estimate of the average effect of this randomized information by comparing energy consumption in the households that did and did not receive the information. However, rebate uptake is likely to be small, because only about 5% of households in our sample applied for rebates each year during the study period. For a RED, incremental uptake from randomly distributed information is likely to be a small fraction of this. Thus, such a study would likely require a very large sample size to achieve a statistically significant estimate of what would likely be a very small reduction in average electricity consumption across the treated population. Any attempt to estimate the average treatment effect on the treated, that is, the energy savings for households that received additional information and applied for an energy efficiency rebate, would be plagued by the same lack of a clearly defined counterfactual as our quasi-experimental approach. Researchers would likely not know who in the control population had purchased new appliances, and even if they did, many of the same selection bias concerns would still be present.

Furthermore, even with perfect evaluation of the short-term direct household-level energy effect of an energy efficiency intervention, this would not give a complete picture of the net effect on energy consumption and greenhouse gas emissions. Indirect rebound effects account for potential increases in energy consumption and greenhouse gas emissions due to both potential increases in overall demand for an energy service, because it becomes more efficient and often cheaper, and due to the embodied energy

and emissions associated with the products and services purchased with money saved through improved efficiency. Estimates of indirect rebound effects vary widely depending on the context. Estimates in the 2000s of indirect energy rebound effects from efficiency programs range from -1 to 123% (Lenzen and Dey, 2002; Nässén and Holmberg, 2009; Azevedo, 2014). Estimates of indirect rebound effects in the 2010s range from -57 to 40% (Kratena and Wüger, 2010; Azevedo, 2014), with estimates of indirect greenhouse gas rebound effects ranging from 5 to 17% for electrical efficiency programs (Thomas and Azevedo, 2013). These studies tend to focus on time periods of at most 35 years, making it difficult to project effects beyond that time horizon (Azevedo, 2014). However, projection on decadal timescales has always been prone to large errors, and the importance of gross energy efficiency for greenhouse gas emission reduction will likely decline over time, because the carbon intensity of energy production continues to fall (Schivley et al., 2018; Sherwin et al., 2018). Importantly, a sizeable portion of the literature on indirect rebound effects uses simulation, rather than statistical analysis.

This case illustrates an important principle in the world of big data: When answering a causal question, having a large amount of apparently relevant data is not enough to guarantee a meaningful estimate (or even the right sign), as illustrated in Smith (2020). A larger dataset with more detailed demographic information likely would not have resolved the underlying selection issues in this analysis, even ignoring indirect rebound effects. We are still convinced that observational causal inference has an important role to play for some energy efficiency programs and in many other fields, and again, even an RCT likely would not have resolved the underlying issues in this particular case. However, this story highlights the need for caution in such analyses. Particularly for domains such as residential energy efficiency, which lie at the intersection of engineered systems, public policy, and human behavior, we need to very concretely think through how people respond to economic and policy incentives when deciding what appliances to put in their households and how to use them. Econometrics is well suited to energy efficiency evaluation in instances in which treatment or encouragement can be successfully randomized (Fowlie et al., 2018; Hahn and Metcalfe, 2021), or when an efficiency intervention focuses on an energy service such as central air conditioning that is already present in the home and that a household will not have more than one of (Boomhower and Davis (2020)). This paper illustrates the severe limitations of econometric approaches when evaluating opt-in energy efficiency programs, such as appliance and building efficiency rebates, which cannot be easily randomized and are subject to numerous selection bias and inframarginality issues highlighted above.

9. Accept the Uncertainty?

Most climate change mitigation scenarios require large improvements in energy efficiency, with sustained reductions in the energy intensity of GDP at or above the highest rates ever achieved in the United States (Loftus et al., 2015). In addition, with increasing levels of variable renewable electricity, the timing of electricity consumption becomes ever more important for the cost, reliability, and greenhouse gas emissions and human health impacts of the grid. Thus, measurement of the magnitude and ideally the timing of energy efficiency savings from specific interventions could help prioritize investment in the most cost-effective strategies, thus reducing the cost of addressing climate change.

In cases with a clear counterfactual, econometric evaluation may be able to provide such estimates. In warmer parts of the United States, such as California's Central Valley, over 90% of households already have some form of air conditioning (Palmgren et al., 2010). Thus, rebates for more efficient air conditioners or efficiency improvements (and, for that matter, better insulation) are unlikely to spur new adoption of air conditioning. Boomhower and Davis (2020) produce what we think is a convincing (however, not decisively causal) econometric estimate of the hourly savings from an air conditioner repair program, which happens to align closely with engineering estimates. Such an approach can even capture region-dependent behavioral aspects of air conditioner use that engineering models would be unable to quantify, such as unexpectedly high energy savings at night (Boomhower and Davis, 2020). However, for energy services that may or may not already be present in a home (e.g., clothes washing or drying) or in

cases in which a household may have two or more of a single appliance (e.g., refrigerators), econometric evaluation of rebate programs may not yield improvements over engineering estimates.

California utilities have already reduced the breadth of their rebate offerings from their high point following the 2007–2008 recession, with PG&E now only supporting smart thermostats, high-efficiency heat pump water heaters, and backup generators for well water pumps (PG&E, 2021). This is partially because of a renewed focus on market transformation, which includes rebates to retailers, rather than customers, alongside improved standards and education. This may also be due in part to evidence that a large number of such rebates are inframarginal, rendering the true cost of these programs relatively high compared with other energy efficiency programs (Boomhower and Davis, 2014; Houde and Aldy, 2017).

The lofty goal of precisely estimating seasonal and hourly effects of energy efficiency measures to integrate them directly into electric power resource adequacy planning and renewables integration is probably not possible for the types of energy efficiency rebates studied here, and the same may well be true for many other forms of energy efficiency measures, particularly those with a strong behavioral component.

Still, appliance energy efficiency rebates remain a tool in energy policy makers' tool kits. Engineering estimates suggest that these rebates save energy if they encourage consumers to purchase more efficient appliances than they would otherwise. The availability of these rebates, in addition to efficiency codes and standards, also encourages manufacturers to prioritize energy efficiency improvements, transforming the market. Unfortunately, all of these effects are difficult to quantify with precision beyond engineering estimates, perhaps coupled with market-level econometric evaluations of changes in appliance sales trends (Houde and Aldy, 2017).

In short, rebates for efficient household appliances may have an important role to play in our energy future, but we likely will not be able to precisely determine how much energy is being saved and when. Tracking the presence or absence of a less efficient appliance as a requirement for rebate participation, and requiring recycling in some instances, may assist with econometric evaluation in some cases, as can surveys of rebate applicants related to confounding factors such as simultaneous building modifications and changes in household size. However, such measures can only partially address these uncertainties. To a certain extent, we will have to accept the uncertainty.

Abbreviations

AMI	Advanced metering infrastructure
IPMVP	International Performance Measurement and Verification Protocol
kWh	Kilowatt hour
PG&E	Pacific Gas and Electric Company
PSM	Propensity score matching
RCT	Randomized controlled trial
RED	Randomized encouragement design
TRM	Technical Reference Manual
UMP	Uniform Methods Project

Acknowledgments. Helpful discussions and comments were received from N. Horner, A. Davis, K. Palmer, P. Ferreira, F. Sowell, G. Morgan, S. Matthews, K. Gillingham, and D. Dzombak, and from Pacific Gas and Electric Company (PG&E) employees B. Smith, C. Kerrigan, A. Doeschot, and H. Liu. A. Mathur helped design the household survey. We gratefully acknowledge PG&E and the Wharton Customer Analytics Initiative for providing us with data. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.

Author Contributions. Conceptualization: All authors; Data curation: E.S. and R.M.; Formal analysis: E.S. and R.M.; Funding acquisition: All authors; Investigation: All authors; Methodology: All authors; Project administration: All authors; Resources: I.A.; Software: E.S. and R.M.; Supervision: E.S. and I.A.; Validation: All authors; Data visualization: E.S.; Writing—original draft: All authors; Writing—review & editing: All authors. All authors approved the final submitted draft.

Competing Interests. Evan D. Sherwin was employed as an intern at Pacific Gas & Electric Company (PG&E) in the summer of 2017 on projects not directly related to this work. Russell M. Meyer is employed by Oracle Corporation. Oracle provides energy efficiency program administration and information technology solutions to the utility sector, including to PG&E.

Data Availability Statement. The utility customer data underlying most of this analysis cannot be released for confidentiality reasons. The data and code from the household survey are available on GitHub at <https://github.com/esherwin/residential-efficiency>. To view supplementary material for this article, please visit <http://doi.org/10.1017/eds.2021.1>.

Funding Statement. Support for this work was provided in part by the Steinbrenner Center for Environmental Education at Carnegie Mellon University. This work was also funded in part by the Center for Climate and Energy Decision Making (SES-0949710 and SES-1463492), through a cooperative agreement between the National Science Foundation and Carnegie Mellon University. This material is based upon work supported by the National Science Foundation Graduate Research Fellowship Program under Grant No. DGE-1252522.

References

- ACEEE** (2019) *State and Local Policy Database: Evaluation, Measurement, & Verification*. Washington, DC: American Council for an Energy-Efficient Economy.
- Allcott H and Greenstone M** (2017) Measuring the welfare effects of residential energy efficiency programs. NBER Working Paper 97. National Bureau of Economic Research. Cambridge, MA, USA.
- Allcott H and Kessler JB** (2019) The welfare effects of nudges: A case study of energy use social comparisons. *American Economic Journal: Applied Economics* 11, 236–276. <https://doi.org/10.1257/app.20170328>
- Azevedo IL, Morgan MG, Palmer K and Lave LB** (2013) Reducing U.S. residential energy use and CO₂ emissions: How much, how soon, and at what cost? *Environmental Science & Technology* 47, 2502–2511. <https://doi.org/10.1021/es303688k>
- Azevedo IML** (2014) Consumer end-use energy efficiency and rebound effects. *Annual Review of Environment and Resources* 39, 393–418. <https://doi.org/10.1146/annurev-environ-021913-153558>
- Bajwa M and Cavicchi J** (2017) *Growing Evidence of Increased Frequency of Negative Electricity Prices in U.S. Wholesale Electricity Markets*. IAEE Energy Forum, pp. 37–41.
- Berg W, Vaidyanathan S, Jennings B, Cooper E, Perry C, DiMascio M and Singletary J** (2020) *The 2020 State Energy Efficiency Scorecard*. Washington, DC: American Council for an Energy-Efficient Economy.
- Berger DJ and Ucar DF** (2013) Comparison of pooled and household-level usage impact analysis. *Paper presented at the International Energy Program Evaluation Conference in Chicago, IL*.
- Blumstein C and Stoff SE** (1995) Technical efficiency, production functions and conservation supply curves. *Energy Policy* 23, 765–768. [http://doi.org/10.1016/0301-4215\(95\)00073-R](http://doi.org/10.1016/0301-4215(95)00073-R)
- Boomhower J and Davis L** (2020) Do energy efficiency investments deliver at the right time? *American Economic Journal: Applied Economics* 12, 115–139. <https://doi.org/10.1257/app.20170505>
- Boomhower J and Davis LW** (2014) A credible approach for measuring inframarginal participation in energy efficiency programs. *Journal of Public Economics* 113, 67–79. <https://doi.org/10.1016/j.jpubeco.2014.03.009>
- Borenstein S and Davis LW** (2016) The distributional effects of U.S. clean energy tax credits. *Tax Policy and the Economy* 30, 191–234.
- Brown MA, Levine MD, Romm JP, Rosenfeld AH and Koomey JG** (1998) Engineering economic studies of energy technologies to reduce greenhouse gas emissions: Opportunities and challenges. *Annual Review of Energy and the Environment* 23, 287–385.
- Burlig F, Knittel C, Rapson D, Reguant M and Wolfram C** (2020) Machine Learning from Schools about Energy Efficiency. *Journal of the Association of Environmental and Resource Economists* 7(6), 1181–1217. <https://www.journals.uchicago.edu/doi/full/10.1086/710606>
- Coito F and Rufo M** (2002) *California's Secret Energy Surplus: The Potential for Energy Efficiency*. The Energy Foundation and The Hewlett Foundation. Menlo Park, CA.
- EPA** (2016) *Heating and Cooling Degree Days*. Washington, DC: Environmental Protection Agency.
- EVO** (2012) *International Performance Measurement and Verification Protocol: Concepts and Options for Determining Energy and Water Savings*, Vol. 1. Washington, DC: Efficiency Valuation Organization.
- Fels MF** (1986) PRISM: An introduction. *Energy and Buildings* 9, 5–18.
- Fowle M, Greenstone M and Wolfram C** (2018) Do energy efficiency investments deliver? Evidence from the weatherization assistance program. *The Quarterly Journal of Economics* 133, 1597–1644. <https://doi.org/10.1093/qje/qjy005>
- Goldstein DB** (2008) Extreme efficiency: How far can we go if we really need to? *Paper Presented at the ACEEE Summer Study on Energy Efficiency in Buildings*, Pacific Grove, CA.
- Hahn RW and Metcalfe RD** (2021) Efficiency and equity impacts of energy subsidies. *American Economic Review* 111, 1658–1688. <https://doi.org/10.1257/aer.20180441>
- Houde S and Aldy JE** (2017) Consumers' response to state energy efficient appliance rebate programs. *American Economic Journal: Economic Policy* 9, 227–255. <https://doi.org/10.1257/pol.20140383>
- Jackson T** (1995) Joint implementation and cost-effectiveness under the framework convention on climate change. *Energy Policy* 23, 117–138.

- Koomey JG, Atkinson C, Meier A, McMahon JE, Boghosian S, Atkinson B, Turiel I, Levine MD, Nordman B and Chan P** (1991) *The Potential for Electricity Efficiency Improvements in the US Residential Sector (Nos. LBL-30477 and UC-350)*. Energy Analysis Program, Applied Science Division, Lawrence Berkeley Laboratory, University of California. Berkeley, CA.
- Kratena K and Wüger M** (2010) The full impact of energy efficiency on households' energy demand. WIFO Working Paper 356. Vienna, Austria: Austrian Institute of Economic Research.
- Lenzen M and Dey CJ** (2002) Economic, energy and greenhouse emissions impacts of some consumer choice, technology and government outlay options. *Energy Economics* 24(4), 377–403.
- Levine MD, Koomey JG, Price LK and Martin NC** (1997) *Scenarios of U.S. Carbon Reductions: Potential Impacts of Energy-Efficient and Low-Carbon Technologies by 2010 and Beyond (Nos. ORNL-444 and LBNL-40533)*. Oak Ridge National Laboratory and Lawrence Berkeley National Laboratory. Berkeley, CA.
- Li M and Dietsch N** (2017) *SEE Action Guide for States: Guidance on Establishing and Maintaining Technical Reference Manuals for Energy Efficiency Measures (No. DOE/EE-1610)*. United States Department of Energy. Berkeley, CA.
- Li, M., Haeri, H., Reynolds, A.,** 2017. *The Uniform Methods Project: Methods for Determining Energy-Efficiency Savings for Specific Measures (No. NREL/SR-7A40-68557, 1397159)*. National Renewable Energy Laboratory. <https://doi.org/10.2172/1397159>
- Loftus PJ, Cohen AM, Long JCS and Jenkins JD** (2015) A critical review of global decarbonization scenarios: What do they tell us about feasibility? *WIREs Climate Change* 6, 93–112. <https://doi.org/10.1002/wcc.324>
- McKinsey** (2007) *Reducing the U.S. Greenhouse Gas Emissions: How Much at What Cost? U.S. Greenhouse Gas Abatement Mapping Initiative*. McKinsey & Company. New York, NY.
- McKinsey** (2009) *Pathways to a Low-Carbon Economy Version 2 of the Global Greenhouse Gas Abatement Cost Curve*. McKinsey & Company. New York, NY.
- Meier A, Rosenfeld AH and Wright J** (1982) Supply curves of conserved energy for California's residential sector. *Energy* 7, 347–358.
- Menne MJ, Durre I, Korzeniewski B, McNeal S, Thoma K, Yin X, Anthony S, Ray R, Vose RS, Gleason BE and Houston TG** (2012) Global Historical Climatology Network—Daily, Version 3 [California].
- Meyer, R.M.,** 2014. Analysis of Selected Regulatory Interventions to Improve Energy Efficiency. Carnegie Mellon University, Pittsburgh, PA, USA. <https://www.cmu.edu/ceic/assets/docs/publications/phd-dissertations/2014/russell-meyer-phd-thesis-2014.pdf>
- Nadel S, Shipley A and Elliott RN** (2004) The technical, economic and achievable potential for energy-efficiency in the US—A meta-analysis of recent studies. In *Proceedings of the 2004 ACEEE Summer Study on Energy Efficiency in Buildings*. Pacific Grove, CA. Citeseer, pp. 8–215.
- Nässén J and Holmberg J** (2009) Quantifying the rebound effects of energy efficiency improvements and energy conserving behaviour in Sweden. *Energy Efficiency* 2, 221–231.
- Novan K and Smith A** (2018) The incentive to overinvest in energy efficiency: Evidence from hourly smart-meter data. *Journal of the Association of Environmental and Resource Economists* 5, 577–605. <https://doi.org/10.1086/697050>
- NRC** (2010) *Realistic Prospects for Energy Efficiency in the United States*. National Research Council and National Academies Press. Washington, DC.
- NYSJU** (2019) *New York Standard Approach for Estimating Energy Savings from Energy Efficiency Programs—Residential, Multi-Family, and Commercial/Industrial Measures*. New York: New York State Joint Utilities.
- Palmgren C, Stevens N, Goldberg M and Rothkin K** (2010) *2009 California Residential Appliance Saturation Study (No. CEC-200-2010-004-ES)*. California Energy Commission. Sacramento, CA.
- PG&E** (2017) *PG&E Energy-Efficiency Rebates for Your Home*. Pacific Gas and Electric Company. San Francisco, CA.
- PG&E** (2019) *Take Control with Time-of-Use Rate Plans*. Pacific Gas and Electric Company. San Francisco, CA.
- PG&E** (2021) *2021 Residential Rebates Catalog*. San Francisco, CA: Pacific Gas and Electric Company. San Francisco, CA.
- Qiu Y and Kahn ME** (2018) Better sustainability assessment of green buildings with high-frequency data. *Nature Sustainability* 1, 642–649. <https://doi.org/10.1038/s41893-018-0169-y>
- Recurve** (2020) *Markets Need Transparent Standard Weights and Measures*. Recurve, Inc. San Francisco, CA.
- Richter B, Goldston D, Crabtree G, Glicksman L, Goldstein D, Greene D, Kammen D, Levine M, Lubell M, Savitz M, Sperling D, Schlachter F, Scofield J, and Dawson J** (2008) How America can look within to achieve energy security and reduce global warming. *Reviews of Modern Physics* 80, S1.
- Rosenfeld A, Atkinson C, Koomey J, Meier A, Mowris RJ and Price L** (1991) *Compilation of Supply Curves of Conserved Energy (No. LBL-31700)*. Berkeley, CA: Center for Building Science, Lawrence Berkeley Laboratory.
- Rosenfeld A, Atkinson C, Koomey J, Meier A, Mowris RJ and Price L** (1993) Conserved energy supply curves for U.S. buildings. *Contemporary Economic Policy* 11, 45–68.
- Rosenfeld AH** (1999) The art of energy efficiency: Protecting the environment with better technology. *Annual Review of Energy and the Environment* 24, 33–82.
- Rubin ES, Cooper RN, Frosch RA, Lee TH, Marland G, Rosenfeld AH and Stine DD** (1992) Realistic mitigation options for global warming. *Science* 257, 148. <https://doi.org/10.1126/science.257.5067.148>
- Saunders HD, Roy J, Azevedo IML, Chakravarty D, Dasgupta S, de la Rue du Can S, Druckman A, Fouquet R, Grubb M, Lin B, Lowe R, Madlener R, McCoy DM, Mundaca L, Oreszczyn T, Sorrell S, Stern D, Tanaka K and Wei T** (2021) Energy

- efficiency: What has research delivered in the last 40 years? *Annual Review of Environment and Resources* 46, 135–165. <https://doi.org/10.1146/annurev-environ-012320-084937>
- Schivley G, Azevedo I and Samaras C** (2018) Assessing the evolution of power sector carbon intensity in the United States. *Environmental Research Letters* 13, 064018. <https://doi.org/10.1088/1748-9326/aabe9d>
- Sherwin ED and Azevedo IML** (2020) Characterizing the association between low-income electric subsidies and the intra-day timing of electricity consumption. *Environmental Research Letters* 15, 094089. <https://doi.org/10.1088/1748-9326/aba030>
- Sherwin ED, Henrion M and Azevedo IML** (2018) Estimation of the year-on-year volatility and the unpredictability of the United States energy system. *Nature Energy* 3, 341–346. <https://doi.org/10.1038/s41560-018-0121-4>
- Smith G** (2020) Data mining fool's gold. *Journal of Information Technology* 35, 182–194. <https://doi.org/10.1177/0268396220915600>
- Sweeney JL** (2016) *Energy Efficiency: Building a Clean, Secure Economy*. Stanford, CA: Hoover Institution Press.
- Thomas BA and Azevedo IL** (2013) Estimating direct and indirect rebound effects for U.S. households with input–output analysis. Part 2: Simulation. *Ecological Economics* 86, 188–198.
- Ürge-Vorsatz D, Novikova A, Köppel S and Boza-Kiss B** (2009) Bottom-up assessment of potentials and costs of CO₂ emission mitigation in the buildings sector: Insights into the missing elements. *Energy Efficiency* 2, 293–316. <https://doi.org/10.1007/s12053-009-9051-0>

Cite this article: Sherwin ED, Meyer RM, and Azevedo IML, (2022). Limitations of econometric evaluation of nonrandomized residential energy efficiency programs: A case study of Northern California rebate programs. *Environmental Data Science*, 1: e1. doi:10.1017/eds.2021.1