Thomas J. Kniesner* and Galib Rustamov

Differential and Distributional Effects of Energy Efficiency Surveys: Evidence from Electricity Consumption[1]

Abstract: Our research investigates the effects of residential energy efficiency audit programs on subsequent household electricity consumption. Here there is a one-time interaction between households, which participate voluntarily, and the surveyors. Our research objective is to determine whether and to what extent the surveys lead to behavioral changes. We then examine how persistent the intervention is over time and whether the effects decay or intensify. The main evaluation problem here is survey participants' self-selection, which we address econometrically via several non-parametric estimators involving kernel-based propensity-score matching. In the first method we use difference-in-differences (DID) estimation. Our second estimator is quantile DID, which produces estimates on distributions. The comparison group consists of households who were not yet participating in the survey but participated later. Our evidence is that the customers who participated in the survey reduced their electricity consumption by about 7%, on average compared to customers who had not yet participated in the survey. Considering the total number of high-usage households participating in the survey in 2009, we estimate that electricity consumption was reduced by an aggregate of 2 million kWh per year, which is approximately equal to the monthly consumption of 3500 typical households in California with an estimated 1527 metric tons less of carbon dioxide emissions. Because the energy audit program is inexpensive ($10–$20 per household) a key issue is that while the program is cost-effective, is it regressive? We find that as the

*Corresponding author: Thomas J. Kniesner, Claremont Graduate University, USA
and
Syracuse University (Emeritus), USA
and
IZA, e-mail: tom.kniesner@cgu.edu
Galib Rustamov: Claremont Graduate University, USA

quantiles of the outcome distribution increase, high-use households save proportionally less electricity than do low-use customers. Overall, our results imply that program designers can better target low-use and low-income households, because they are more likely to benefit from the programs through energy savings.

Keywords: difference-in-differences;electricity consumption;information salience; multiple treatments; propensity-score matching; quantile regression; self-selection; treatment effects.

JEL classifications: C31; D03; D12; L94; Q41.

1 Introduction

Home energy audits have been offered in the United States since the 1970s, and the use of the audits has expanded with the availability of stimulus funds in recent years (Ingle et al., 2012). In California, home energy efficiency survey (HEES) programs are implemented statewide by public utilities. The programs' objectives are to increase awareness, inform customers about their consumption behavior, and make other resources available to reduce energy consumption. When customers complete a survey questionnaire they receive extensive personalized feedback and tips about what actions they can take to save energy and money. The surveys inform both the implementer and the consumer how energy has been used in a house. Because there is imperfect information regarding a household's inattention and usage behavior, personalized feedback can lead to the desired behavioral change. We present evidence that customers who participated in the survey reduced their electricity consumption by about 7% on average compared to customers who had not yet participated in the survey, and as the quantiles of the outcome distribution increase, the effect of the program decreases.

As discussed in earlier studies, consumers may behave inefficiently because of the unclear relationship between price and electricity use. Home energy efficiency audits can close the information gap via personalized feedback serving as a reminder. By providing additional tailored information, personalized feedback may also decrease information asymmetry and result in more efficient and persistent behavior change by lowering the cognitive cost of energy decision-making (Gillingham et al., 2009). Numerous conservation studies have been designed using varying informational and behavioral strategies to address to the information gap (Abrahamse et al., 2005; Delmas et al., 2013). Utility companies have been one of the major implementers of home energy audit programs. Under regulatory practices utilities have an incentive to invest in conservation measures, but they may

limit actual conservation through the improper design of a program (Wirl & Orasch, 1998).[2] So, although we may encounter some successes the ways that home energy audits have been designed and executed have often been ineffective.

In California, utilities also have not used research methods (RCT design) to build and implement HEES programs. Thus, this and similar types of self-selected participant studies did not lead to ground-breaking policy changes or the behavioral interventions needed to change consumer behavior. Recently, there have been some signs of the implementation of scientific approaches in energy efficiency program designs. In California the Public Utilities Commission (CPUC) made it mandatory for all statewide Investor-Owned Utilities (IOUs) to implement behavior-based programs. An example of such a behavior-based program is the implementation of social comparisons by the research company OPOWER (Allcott, 2011) or multi-family complex competition by the Southern California Edison (SCE) Company (Chen et al., 2015).

Much of the empirical microeconomic literature in the economic development area uses econometric and statistical methods to overcome the deficiencies of non-experimental data (Deaton, 2000). Because of the inherent self-selection in the survey we study, we begin by employing the empirical technique of Sianesi (2004), who examines the effectiveness of unemployment programs in Sweden. She suggests selecting future program participants for matching estimations. We apply the method in a different market setting, residential energy efficiency audits. The DID estimator provides evidence that participation in the survey leads to about 7% less electricity consumption by survey participants on average compared to customers who did not participate. In addition, the effect is persistent over time, at least for the year after the survey.

Our objective here is to propose an alternative method when evaluating HEES programs by selecting future participants as the control group as suggested by Sianesi (2004). The approach is different from the current practice of evaluating HEES and similar programs. For instance, 2006–2008 HEES impact evaluation was based on the participant information only, whereas the 2010–2012, and 2015 studies were based on using non-participants to match and select as the control group to estimate the treatment effects.[3]

---

2 All investor-owned electric and gas utilities in California engage in decoupling. Decoupling separates electricity retailers' profits from quantities sold and is one mechanism that could encourage firms to nudge consumers toward reducing energy usage (Brennan, 2010; Allcott & Mullainathan, 2010). Specifically, decoupling does not provide an affirmative incentive for utilities to encourage conservation; it simply removes a disincentive not to conserve. Thus, utilities have not had strong incentives to provide efficient ways to implement home energy audits. There is also a cost associated with the implementation of effective tools to change the behavior of the majority of the utilities' customer bases.
3 For further discussion regarding the impact evaluations of the previous program cycles, please see both Itron (2013) and Cadmus (2017) evaluation reports.

Utilities use various delivery mechanisms to implement home energy audit programs – through mail, online, telephone and in-home (on-site) audits. Here we also investigate the differential performance of the mail-in versus online versions of the home energy surveys in addition to the combined survey impact.[4]

Finally, there is recent concern as to whether nudge-based and other household energy conservation programs are regressive and whether onerous requirements are imposed on less well-off households (Gayer & Viscusi, 2013; Levinson, 2016). We employ quantile regression techniques to detect the distributional usage effects of the home energy surveys. Households in the lowest quantiles have more substantial response to non-binding energy conservation efforts, in percentage reduction of electricity consumption, than consumers in the median or highest quantiles. The importance of our quantile analysis is in showing that the estimated survey effects differ by the level of pre-survey household consumption.

2 Data

The data that an IOU in California provided to us are on a confidential basis. The information covers more than 4200 customers who voluntarily participated in the HEES in January of 2009 and 2010. We eliminated households with less than 12 months of consumption data during the period, leaving a total of $N = 4173$ households.

Because households opt-in to the HEES program we first chose the January 2009 survey participants as the treatment group and the future survey participants, those from January 2010, as the comparison group. The comparison group contains customers who did not participate in January 2009 and have not yet participated in the survey (Sianesi, 2004). We use same monthly usage and billing interval, 2008 and 2009, for treatment and comparison groups. The summary statistics for our data appear in Table 1.[5] The data set we use here is the result of combining three main sources that reflect monthly energy consumption: billing, dwelling demographics, temperature and the survey (HEES). The billing data cover 2008 and 2009 for both the 2009 and 2010 survey participants. The weather information comes from the monthly Cooling Degree-Days (CDD) data over the billing period from 2008 to 2009, which we merged with the main dataset. Because California has warmer

---

4 Because of fewer observations we excluded on-site and telephone surveys. For instance, during the 2010–2012 program cycle, on-site and telephone surveys made up close to one percent of the participants (Itron, 2013).

5 All models control for a household billing, demographics, dwelling characteristics, survey data as well as weather variables.

Table 1 Summary statistics, residential accounts and energy usage.

		All	**January 2009 participants**	**January 2010 participants**
Observations		99,568	21%	79%
	Mail-in	82,566	15%	85%
	Online	17,002	53%	47%
Number of accounts		4173		
Mean Log Consumption (kWh)		7.03 (0.51)	6.86 (0.71)	7.07 (0.43)
	Mail-in	7.10 (0.42)	7.09 (0.62)	7.11 (0.39)
	Online	6.67 (0.70)	6.89 (0.71)	6.78 (0.65)

Note: Standard deviations are in parentheses. Percentages are rounded. 97.5% of the households in the data have 24 months of observation, 2.5% varies between 15 and 23 months.

weather than the national average, we used a 72 °F indoor baseline temperature instead of the nationally defined baseline of 65 °F.

The HEES program provides residential customers with an energy audit of their homes through a mail-in, online, telephone, or in-home (on-site) energy survey. The survey instrument asks the participants a series of questions about their homes and then offers a list of tips based on their responses. Subsequent recommendations include both possible changes in behavior and information on more energy-efficient appliances. The program is meant to incite action; its purpose is to inform the participants of opportunities to save money and to provide resources to implement the recommendations.

It is important to determine whether the design of the HEES report is successfully imparting useful knowledge, referring participants to helpful resources, and whether the coordination effort is motivating participants to adopt more energy- and water-efficient behaviors. As noted earlier we focus on mail-in and online survey participant data. The two survey methods are commonly compared with other methods. Furthermore, telephone and in-home surveys are being used less frequently by utilities and have not been the most preferred choices of participation by the customers.[6] In-home data are also costly for utilities to collect, although the largest savings are observed as a result of an in-home survey-based intervention (ECONorthwest, 2009; Itron Inc., 2013).

The literature presents evidence of low take-up rates to opt-in energy efficiency programs and home energy reports (HERs) (Fowlie et al., 2015; Allcott & Kessler, 2018). Throughout California IOUs have used various targeting methods to get

---

6 See 2010–2012 CPUC HEES Impact Evaluation by Itron (2013).

customers to complete the energy efficiency surveys. The marketing process often targeted households with high bills and therefore likely to achieve higher savings, which is particularly true for mail-in participants in our sample (Itron, 2013).[7] To encourage households to complete the surveys, IOUs provided incentives, such as gift cards, to participate in the survey program (Itron, 2013). Online surveys also marketed with email blasts and through utility websites, but online surveys were still available to all households through the IOU website.[8]

Figure 1 describes the difference among the treatment, comparison, and randomly selected households in terms of their monthly average electricity usage. It highlights the mean energy usage differences (kWh) by income groups during the pre-survey period. Randomly selected households, who never participated the HEES program, total about 10,000 residential utility households from the same utility company.[9] We show means for each income groups by treatment versus comparison sub-groups. Randomly selected non-participant households consumed substantially less energy than households in both the treatment and comparison groups, who participated the survey in the following year (2010). Overall, HEES participants who opted into the program had a higher average usage than non-participants.
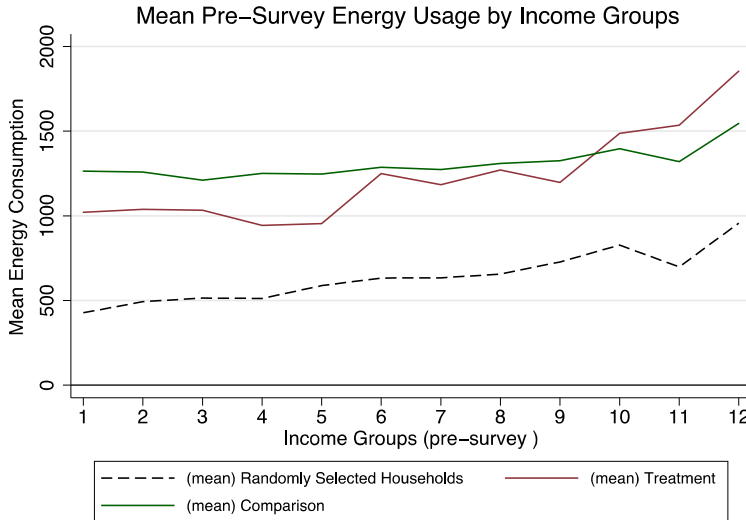
3 Methods

Because the audit program uses online-based and mail-delivery mechanisms (formats) to reach customers, we first evaluate the average impact of each format separately on post-audit energy consumption behavior. Here the treatment group is the January 2009 program participants, and the comparison group is the January 2010 program participants. To address the self-selection issue we first identify a valid comparison group. We chose the January 2010 program participants (future survey participants) as the comparison group so that the classical treatment and control

---

7 Post-survey interviews with some households across all IOUs reported that saving money and high-energy bills were the largest motivational factors for taking the HEES survey (Itron, 2013). In addition, the Itron (2013) report shows that the majority of the interviewees also knew about the online survey service, but preferred to take the mail-in survey due to convenience (61%), reluctance to share personal information online, and internet access (20%). The primary reason to participate in the surveys (saving money, lower bills) is reported to be only 53% of the households' motivation. Households show a diverse set of reasoning about their motivation to complete the survey and receive personalized feedback. For more detailed discussion about the program, see http://calmac.org/results.asp?t=2.

8 IOUs often provide process evaluation reports for each program cycle, and during each program cycle new customers often being targeted to participate the program. These reports are posted on California Measurement Advisory Council (CALMAC) website, and publicly available.

9 Additionally, average monthly residential electricity usage in California is about 573 kWh (http://www.electricitylocal.com/states/california/).

Mean Pre–Survey Energy Usage by Income Groups

Figure showing mean energy consumption by income groups for three groups.

Note: Randomly selected households are those who never participated in the survey. The data for this group are consist of about 10,000 residential utility households from the same IOU, who did not participated the survey during the survey period. Incomes are bracketed as following in the IOU data:
1.A=$1 - $15,000; 2.B=$15,000 - $24,999; 3.C=$25,000 - $34,999; 4.D=$35,000 - $49,999; 5.E=$50,000 - $74,999; 6.F=$75,000 - $99,999; 7.G=$100,000 - $124,999; 8.H=$125,000 - $149,999; 9.I=$150,000 - $174,999; 10.J=$175,000 - $199,999; 11.K=$200,000 - $249,999; 12.L=$250,000+

Figure 1 Following graph shows the mean energy usages (kWh) by income groups during the pre-survey period. Non-participant households consumed substantially less energy than households in both treatment and comparison groups. In our survey sample, only 13.4% of observation of the survey participants is below $50,000.

distinction holds (Sianesi, 2004). Our framework then determines the proper and valid matching estimation. The approach we use is more reliable (Sianesi, 2004, 2008) than matching persons who have never participated in home energy audits (Du et al., 2014; Itron, 2013). The HEES program evaluation study prepared by Itron, Inc. (2013) also presents the impact of the survey by employing a matching method where the comparison group also is non-participants.[10]

Another common practice in evaluating home energy audits has been engineering-based ex ante analysis, which has led to systematically biased and exaggerated energy savings estimates and significant overestimates of persistent energy saving (Nadel & Keating, 1991; Dubin et al., 1986; Davis et al., 2014;

---

10 Impact evaluation for 2009 HEES participants was not available, so we cite and compare our approach to the method and empirical design of the HEES program evaluations of 2010–2012 and 2006–2008 program cycles.

Gerarden et al., 2017). In particular, "There may have been a selection bias whereby researchers have chosen to evaluate engineering-economic analysis that have most exaggerated the savings potential of efficiency investments" (Gerarden et al., 2017).

3.1 Addressing self-selection in opt-in programs

Randomized experiments create independence between the treatment application and consumer characteristics, both observed and unobserved. Non-randomized observational data can be misleading because of self-selection – decisions made here by households to participate in the energy efficiency survey. The main concerns are unmeasured factors, such as motivation to take action, which may affect the decision to participate in the survey along with post-intervention behavior. A customer who has requested an audit may be from the type of household taking other unobserved actions to conserve energy (Allcott & Mullainathan, 2010).

The confounding difference between survey participants and non-participants underscores the difficulty of controlling for interpersonal differences when estimating the causal effects of programs. The main problem here is that often the researcher wishes to draw conclusions about the wider population, not just the sub-population from which the data come (Kennedy, 2003). However, because of ethical problems, the large costs of implementing randomizations, and problems with external validity, many studies use observational data instead of implementing a randomized experiment (Fu et al., 2007; Black, 1996).

Similar to many other energy efficiency survey programs the HEES audit program we evaluate here acknowledges that the customer chooses to participate in the survey instead of having been randomly assigned by the program designer. Because people self-select into the program, it complicates identifying what the response will be if the program were implemented on a mandatory basis or through some added participation incentive payment. However, if the research question is simply how do *voluntary* participants in the programs respond then there is no confounding self-selection issue. Although it seems that households opt-in to the program, IOUs targeted the high-energy users through mailings, post-cards, the IOU website, email blasts, incentives, and various other ways to induce high-usage households to join and complete the survey. So, in our sample the customers, who are high-energy users, were particularly targeted and tagged to be part of the program.[11] This means that the HEES program is similar to the Weatherization Assistance Program (WAP) that Fowlie et al. (2015) studied, where the program provides free energy efficiency

---

[11] Akerlof (1978), Allcott et al. (2015).

improvements to low-income households.[12] The design of the HEES audit program also differs from the solar photovoltaic (PV) programs in California where the rate structure and cost of the PV installation, regardless of the tax incentive and rebates, has tilted more affluent households toward participating (Borenstein, 2017).

Because we want our empirical results to be informative on the issue of mandatory implementation, we also consider econometric solutions to the problem of self-selected data. To provide a proper estimate of the treatment effect with observational data we employ the method suggested by Sianesi (2004), where the comparison group is customers who were not yet participating in the survey but participated later. The samples in both treatment and comparison groups received the same type of encouragement or targeted marketing, but at different times. Our empirical approach suggests a method that could reduce possible inflated program effects estimates and provides a credible approach in assessing the underlying causal hypothesis. Because an experimental approach was not feasible for the type of survey used by the institution, we propose a credible empirical method and comparison group. Here we are not only using a selection on observables approach, but we are also using future participants as a comparison group, which addresses the unobservable characteristics issue. So, we are not only suggesting a method that could address possible self-selection problems in evaluating energy-saving programs, but we are also proposing a method that is credible in measuring the effectiveness of energy-saving programs in targeted sample settings. Instead of using randomly selected utility customers as a comparison group and matching them with the treatment group based on observable pre-survey characteristics, we use customers who joined the program later, in January 2010.[13]

3.2 Evaluation approach

Using the mean outcome of untreated individuals $E[Y_0|T=0]$ in non-experimental studies is usually not a good idea because components that determine the treatment decision may also determine the outcome variable of interest (Caliendo & Kopeinig, 2005). This suggests that even if the researcher chooses the best possible candidate for the comparison group their consumption levels will still be different if consumers do not participate in the surveys because of the unobserved

---

12 Within an opt-in WAP, Fowlie et al. (2015) had a randomized encouragement intervention, where to increase the treatment group's probability of program participation there were different recruitment channels and applications assistance.

13 See Heckman et al. (1998) and Imbens and Rubin (2015) for a detailed discussion about the matching and addressing the comparison group in social programs.

counterfactual. We therefore begin by estimating the effect without matching for comparison with the other models. Then we estimate the program effect using the matching methods. To validate the matching procedure for empirical content and external validity, it is important that the following conditions hold: the conditional independence assumption (CIA) and common support (CS). The CIA suggests that given a set of observable characteristics, the distribution of Y_t^0 for customers who participate in the survey in January 2009 is the same as the (observed) distribution of Y_t^0 for customers who wait until January 2010 to participate (Sianesi, 2004):

$$Y_t^0 \perp T | X = x \quad \text{for } t = \text{January 2009; January 2010.} \quad (1)$$

Because we chose a comparison group from the future participants, equation (1) postulates that conditional on X, there is no unobservable heterogeneity left that affects both survey participation and later consumption (Sianesi, 2004, 2008; Caliendo & Kopeinig, 2005), which suggests that the probability distributions of the two groups are similar to each other.[14]

Another requirement for the matching methods procedure is the CS or overlap condition:

$$0 < Pr(T = 1 | X) < 1. \quad (2)$$

"This condition guarantees that persons with the same X values have a positive probability of being both participants and non-participants" (Heckman et al., 1999). The CS condition means that for every customer in the treatment group there are customers with similar characteristics in the comparison group. Heckman et al. (1999) show that the CS condition is central to the validity of matching. Considering the conditional independence and CS conditions, the literature suggests that the propensity score is useful in constructing matching estimators. The propensity score is the conditional probability of being treated at time t given a vector of observed characteristics, which reduces the dimensionality of the matching problem (Rosenbaum & Rubin, 1983). The propensity score here estimates the propensity of the customers with a set of observed characteristics to receive the program – the energy efficiency survey.[15] Thus, the customers who have the same or similar propensity-score values have similar distributions of all of the observable characteristics.

---

14 We use the pre-treatment characteristics of X for the CIA.

15 We use the conditional propensity score based on pre-treatment observable characteristics such as income, weather (CDD), house ownership and type of house. The idea is to find "lower-dimensional functions of the covariates that suffice for removing the bias associated with the differences in the pre-treatment variables" (Imbens & Rubin, 2015). It is both difficult and not efficient to employ a large number of covariates. Considering both graphical and empirical results we will see that the estimated conditional propensity score was appropriate for continuing to calculate the non-parametric estimators of interest here.

Figure 2 shows that the customers in the treatment and comparison groups have similar propensity-score distributions. According to Dehejia and Wahba (2002), propensity-score-matching estimates are more consistent with estimates that are derived from an experimental design. However, propensity-score matching does not guarantee that all of the individuals in the non-treatment group will be matched with individuals in the treatment group (Titus, 2007).

Once estimated, the propensity score can be used in a variety of analytic approaches, such as matching and weighting. The literature identifies several ways of matching each survey participant to a non-participant (Rosenbaum & Rubin, 1983,1985; Rubin & Thomas, 1992; Baser, 2006; Hansen, 2004; Smith, 1997). We use kernel propensity-score-matching methods to calculate the difference-in-differences estimator. Kernel matching is a non-parametric estimator that uses weighted averages of all persons in the comparison group to construct the counter-factual outcome (Caliendo & Kopeinig, 2005). The kernel-based weight declines with the distance between the individuals in the two groups. No specific matching estimator is appropriate by itself. We performed kernel-based propensity-score matching because of the large sample size and feasibility.

We then introduce non-parametric versions of the difference-in-differences (DID) estimation with the later participants as a comparison group using the kernel-based propensity-score-matching method (Meyer, 1995; Heckman et al., 1998; Sianesi, 2004, 2008; Allcott, 2011). Allcott (2011) suggests forming a comparison group according to the average monthly energy use of households. The benefit of the standard DID model is that it provides the average effect of the intervention on the treatment. Furthermore, because of the self-selection in the sample, we adopt a difference-in-differences matching estimator to control for the presence of the unobservable characteristics, as referenced in List et al. (2003). Finally, Heckman et al. (1998) and Blundell and Costa Dias (2009) note that propensity-score DID accounts for both observed and unobserved time-invariant differences between the treatment and the comparison groups, which mitigates bias.

The design of our DID model is as follows. Individual i belongs to either the treatment or the comparison group, $T_i \in \{0, 1\}$, where $T = 1$ is the treatment group and is observed in t periods, where t indexes to periods 1 and 2. The period of i's consumption behavior is defined as $P_t \in \{0, 1\}$, before and after treatment periods. Y_i is the outcome variable – monthly energy consumption in ln(kWh) and in kWh. The interaction term $T_i \cdot P_t$ is an indicator of the treatment. The standard DID model for the realized outcome is then

$$Y_{it} = \alpha + \beta T_i + \gamma P_t + \psi(T_i \cdot P_t) + \theta X_{it} + \epsilon_{it}. \tag{3}$$

The coefficient of the interaction term, ψ, is the DID effect, or the impact of survey participation on later consumption behavior. X is a vector of household
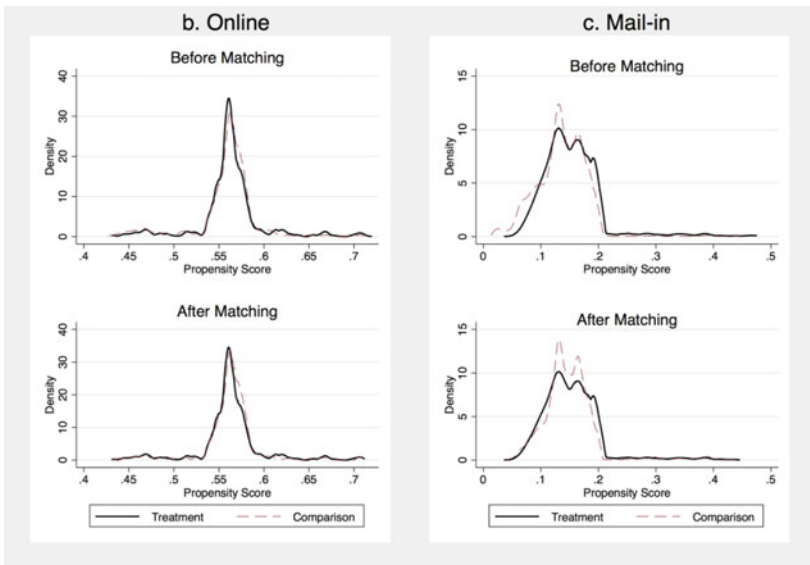
a)
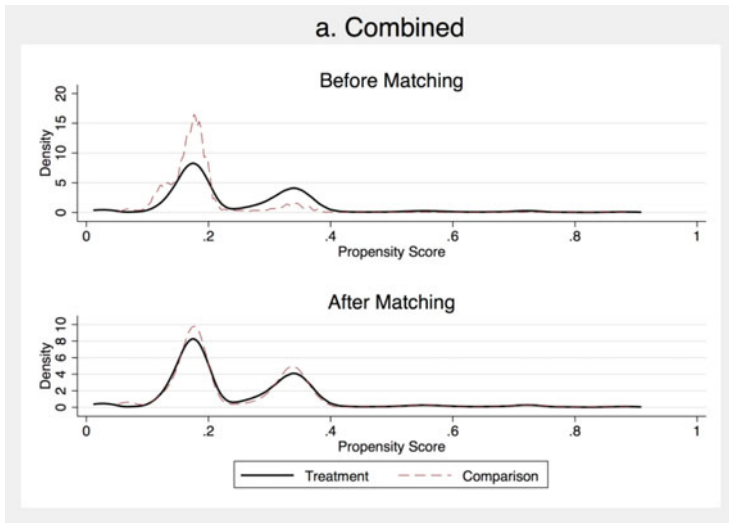
b)

Figure 2 Figure shows estimated propensity of scores, by groups – treatment and comparison. Pre- and post-matching density estimates of propensity scores among the treatment and comparison groups (Epanechnikov kernel, the bandwidth is 0.06 – default).

demographics, dwelling characteristics and responses to the survey questionnaire. The DID is the difference in the average outcome in the treatment group before and after the treatment minus the difference in the average outcome in the comparison group before and after the treatment (Athey & Imbens, 2006). Following equation shows the standard DID estimand, ψ.

$$\psi^{DID} = [\mathbb{E}[Y_{it}|T_i = 1, P_t = 1, X_i] - \mathbb{E}[Y_{it}|T_i = 1, P_t = 0, X_i]]$$
$$- [\mathbb{E}[Y_{it}|T_i = 0, P_t = 1, X_i] - \mathbb{E}[Y_{it}|T_i = 0, P_t = 0, X_i]]. \quad (4)$$

Smith and Todd (2001), who examine whether social programs can be reliably evaluated without using randomized experiments, conclude that DID matching estimators generally exhibit better overall performance. Considering that our study has access to the pre- and post-treatment residential energy consumption data, DID with propensity-score-matching approach is suitable for our research.

Another type of non-parametric approach that we apply is the quantile DID (QDID) matching method. We continue using kernel-based propensity-score matching. The focus for the basic DID method is to produce the average causal effects of program participation. However, we are also interested in investigating the effect of the programs on the entire distribution of outcomes. "The distribution of the dependent variable may change in many ways that are not revealed or are only incompletely revealed by an examination of averages" (Frölich & Melly, 2010). Because our dependent variable is continuous – monthly energy consumption – it makes sense to test the effect on the distribution by identifying the relative savers and losers (Angrist & Pischke, 2009). The primary observable source of heterogeneity is as a function of pre-treatment usage (Allcott, 2011). It is possible that households in the lower quantiles respond to the survey differently than households in the upper quantiles. Quantile regression reduces the importance of outliers and functional-form assumptions and allows us to examine features of the distribution besides the mean (Meyer et al., 1995).

Here the survey may have different effects in different quantiles, so that we apply DID to each quantile rather than to the mean to investigate features of the distribution (Meyer et al., 1995; Athey & Imbens, 2006). The QDID estimates we present are for both the extreme (0.1 and 0.9) and central (0.25, 0.5, 0.75) quantiles. The QDID estimator on quantile q can be written as

$$\psi_q^{QDID} = F_{Y,11}^{-1}(q|X) - F_{Y,10}^{-1}(q|X) - [F_{Y,01}^{-1}(q|X) - F_{Y,00}^{-1}(q|X)], \quad (5)$$

where $F_Y^{-1}(q|X)$ is the distribution function for Y at q, which is conditional on X (the matched observable characteristics or propensity scores) (Athey & Imbens, 2006). Equation (5) shows the difference between treatment and comparison groups before and after the treatment for different quantiles. To our knowledge, our study

is one of the earliest attempts to apply the QDID matching method to residential energy efficiency program evaluation.

We use the natural logarithmic transformation, Ln(kWh), where the interpretation of the effect is in terms of proportionate changes.[16] We show changes in kWh usage as well. Finally, to identify the durability of the intervention we estimate both short-term (quarterly) and longer-term (year) effects of energy efficiency survey participation.

4 Results

Participation in the energy audit program is voluntary. If non-participants are used as a comparison group, systematic energy use differences would be apparent between the participant and non-participant groups because of unobservable motivation and observable household characteristics.[17] In contrast, the Itron (2013) report for CPUC employed propensity-score matching with non-participants and matched with too many observable characteristics. As a result, almost 90% of the samples in the comparison group were dropped during the matching process. We instead begin by focusing on identifying and justifying the valid comparison group and then continue with the regression estimation. The objective is to prevent an inflated estimate of the audit program's potential impact. The interest in calculating the propensity score and matching methods "purely lies in their combined ability to balance the characteristics of the matched sub-groups being pair-wisely compared" (Sianesi, 2008).

We estimate the outcome of interest, post-audit behavior, by employing two non-parametric estimation techniques. We begin with kernel propensity-score-matching DID, which produces average treatment effects. We also investigate the impact of an audit on the entire distribution by employing a QDID approach. Although we focus on overall survey participation, we also report the results separately for web-based and mail-in program participants and the impact on consumption over time. Our results suggest that there is a significant reduction in

---

16 To examine the validity and to verify the results, we calculated the bootstrapped confidence intervals (Lechner, 2002; Black & Smith, 2004; Sianesi, 2004). This can improve the validity of the analysis and mitigate the potential bias of the estimation.

17 We initially began by taking randomly selected customers (non-participants) as a comparison group. The confounding difference between the treatment and comparison groups is sufficiently convincing so as not to pursue the randomly selected non-participants option when we can choose customers who waited longer (one year) to participate in the program. For example, the mean kWh usage among the survey participants is much greater than that of randomly selected residential non-participants (see Figure 1).

consumption overall with audit participation. Web-based survey participants show much greater reactions to their surveys than mail-in participants do (11% vs. 4%).[18] To test the durability of the intervention, we also estimate short-term (quarterly) and longer-term (year) effects. Because we use DID and QDID, seasonality should not be a concern. However, for an additional robustness check, we calculated the estimators in both scenarios – seasonally adjusted and unadjusted regressions – and there is only an incremental difference between the two estimators. The details of the additional analyses and discussions appear in the following sub-sections.

4.1 Graphical results and balance diagnostics

Figure 1 shows the mean energy usages (kWh) by each of the income groups during the pre-survey period. The matching procedure was effective in creating a group of customers who were comparable with the treatment group based on observable confounders. So, first we estimate the probability of participating in the survey given the values of potential confounders (the propensity score) for each customer in the data. Next, we graphically display the distribution of propensity scores of the treatment and control groups (Figures 2a and 2b) for all cases – overall, online and mail-based delivery mechanisms. The graphs show that the distributions of the propensity scores significantly overlap. A visual examination of the before-matching distribution also allows checking of the region of CS. In each graph there is sufficient overlap between the treatment and control groups, which suggests that one can make reasonable comparisons. Then we match individuals in the treatment group with individuals in the comparison group based on the kernel-based propensity scores. Figures 2a and 2b compare the propensity-score distribution of the treatment and comparison groups before and after matching. The density plot graph shows that the propensity scores have similar trends, and the graph reveals an extensive overlap of the distributions.

   Next, we check the balance diagnostics (Table 2). "In the context of propensity-score matching, balance diagnostics enable applied researchers to assess whether the propensity-score model has been adequately specified" (Austin, 2009). Table 2 reports both the bias and the mean differences between the treatment and comparison groups in the matched sample. The matched groups' balance is off by only a

---

18 We also estimate whether differences in reactions to the survey between web and mail participants are statistically significant via the "difference-in-difference-in-differences" (triple difference) method suggested by Hamermesh and Trejo (2000). The triple difference estimation shows the statistical significance of the difference (not shown). Differences in response rates could also be attributed to some differences in household characteristics between mail-in and online survey participants. See the Appendix for elaboration.

Table 2 Balance diagnostics across all the estimated propensity scores.

Variable	Unmatched Matched	Mean		% bias	*t*-test	
		Treatment	Comparison		*t*	*p* > |*t*|
Pscore (Overall)	U	0.25255	0.18823	58.8	85.69	0.000
	M	0.25255	0.25202	0.5	0.39	0.694
Pscore (Online)	U	0.56259	0.55768	14.1	8.64	0.000
	M	0.56246	0.56258	−0.3	−0.24	0.810
Pscore (Mail-in)	U	0.15262	0.13556	35.6	35.46	0.000
	M	0.15172	0.14892	5.8	4.1	0.000

Note: Propensity scores are estimated conditional on pre-treatment (survey) observable characteristics.

small amount, where the value of the standardized bias for overall HEES participation is 0.5%, which is less than the unmatched maximum of about 59%. Moreover, the differences between the groups became statistically insignificant during the post-matching period ($t = 0.39$).

Table 2 also shows the assessments of online and mail-in survey participation. The pre- and post-matching trends for the overall survey and the online survey are close to each other. The standardized bias for online participants is 0.3%, which is also less than the unmatched maximum of about 14%, and which suggests that (even before any matching) the group of online participants was more similar than were the general survey and mail-in participants. In both the overall and online scenarios the propensity score is balanced in the matched sample. In contrast, the pre- and post-matching differences are significant for the mail-in audit participants, and there is a significant reduction in percentage bias: the pre-matching bias was reduced from about 36% to about 6%. Studies suggest that the standardized bias should be less than 5% to 10% (Rosenbaum & Rubin, 1985; Austin, 2009). In addition, the sample size in the data influences the *t*-test (Austin, 2009). For mail-in participants, the number of future participants is much greater than the treatment group, so one should not place undo emphasis on the *t*-test versus on the standardized percent bias.

4.2 Estimation results

We now examine various measures over a two-year period to investigate how customers who participated in the energy efficiency audits performed, on average (individual and distribution), compared to customers who waited one year to participate.

Table 3a The following results show the coefficients of the DID estimator for both standard unmatched (1, 2, and 3) and propensity-score-matching DID (4, 5, and 6) regressions. Dependent variable: Log(kWh) Consumption.

Variables	(1) Ln(kWh)	(2) Ln(kWh)	(3) Ln(kWh)	(4) Ln(kWh)	(5) Ln(kWh)	(6) Ln(kWh)
Diff-in-Diff – A	−0.0572***			−0.0669***		
	(0.00678)			(0.00653)		
Diff-in-Diff – M		−0.0375***			−0.0438***	
		(0.00721)			(0.00591)	
Diff-in-Diff – O			−0.112***			−0.110***
			(0.0178)			(0.0175)
Observations	83,836	70,445	15,157	83,836	69,514	15,131
R^2	0.393	0.322	0.365	0.358	0.360	0.371

Note: Standard errors are in parentheses. Estimations are adjusted for seasonality using seasonal dummies. Ln (kWh): Log Consumption (kWh). A – aggregate, M – Mail, O – Online. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

We begin by presenting the standard DID estimates where the comparison group is not matched based on the kernel propensity-score matching.

Table 3a summarizes the DID estimates, where the outcome is the natural log of electricity consumption. Columns 1–3 show the DID results without matching, and columns 4–6 show the propensity-score estimates. The significance of the coefficients, the small differences among the coefficients (approximately one percent), and the standard errors between the matched and unmatched estimations further verify the validity of the comparison group. Table 3b depicts the same evidence where the dependent variable is kWh consumption. The results in Tables 3a and 3b suggest that one year after an energy audit program participation the customers who participated in the survey in January 2009 reduced their electricity consumption by about 7%, or 76 kWh on average, compared to households that did not participate in the survey until January 2010. Our mean results are consistent with the meta-study of informational conservation experiments, which finds a weighted average effect of about a 7% electricity reduction (Delmas et al., 2013).

The different performance of online survey participation compared to mail-in survey participation is also important. Tables 3a and 3b show that, on average, one year after HEES participation the online HEES participants reduced their electricity consumption more than the mail-in participants, 11% vs. 4% (112 kWh vs. 52 kWh). Du et al. (2014) also report a similar differential effect between online and mail-in HEES participants. In particular, they investigated the probability of

Table 3b The following results show the coefficients of the DID estimator for both standard unmatched (1, 2, and 3) and propensity-score-matching DID (4, 5, and 6) regressions. Dependent variable: kWh Consumption.

Variables	(1) kWh	(2) kWh	(3) kWh	(4) kWh	(5) kWh	(6) kWh
Diff-in-Diff – A	−62.674***			−75.57***		
	(9.278)			(9.259)		
Diff-in-Diff – M		−44.56***			−51.91***	
		(11.26)			(11.16)	
Diff-in-Diff – O			−116.1***			−112.4***
			(17.78)			(17.30)
Observations	83,949	70,454	15,272	83,949	69,526	15,241
R^2	0.2730	0.250	0.350	0.269	0.204	0.362

Note: Standard errors are in parentheses. Estimations are adjusted for seasonality using seasonal dummies.
A – aggregate/combined, M – Mail, O – Online. Matching is based on the kernel-based propensity score.
*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

future energy efficiency program participation as a function of current HEES participation.[19] They conclude that the delivery mechanism of the survey matters for post-intervention behavior. Thus, the households who participated in the online survey increased the probability of future energy efficiency program participation by 3% to 4% compared to under 3% for mail-in survey participation.[20] This suggests that utilities and program designers could achieve greater behavioral responses in terms of reducing electricity consumption or participating in different behavioral programs in the future by promoting online survey mechanisms, which is also the least costly approach.

Table 4a depicts the average treatment effect on later consumption behavior over time. It is important to examine and distinguish the effects of short-term ver-

---

19 The program we study also creates spillover effects beyond reducing energy consumption. According to Du et al. (2014), consumers who participated in HEES programs are also more likely to participate in other behavioral energy efficiency programs in the future. Households that were not responsive to the survey in the short run gradually changed their routines and formed new habits. Although our purpose here is not to examine habit formation exhaustively, habits increase the marginal utility of engaging in an activity in the future (Charness & Gneezy, 2009). Education and information alone will not sufficiently incent a household because of insufficient economic salience in the market and can reverse the effects of policy goals leading to inertia in consumption behavior and investment decisions.
20 In contrast to our empirical approach, Du et al. (2014) select non-HEES participants for matching purposes. Thus, we would expect slightly different results if they match with future HEES participants.

Table 4a Over time Kernel propensity-score-matching DID estimations: Treatment effect of participating the survey on January 2009 compared to waiting until January 2010. ***Combined survey participation.*** Dependent Variable: Log Consumption.

Variables	(1) Ln(kWh)	(2) Ln(kWh)	(3) Ln(kWh)	(4) Ln(kWh)
Diff-in-Diff – A	−0.0122	−0.0304***	−0.0582***	−0.0667***
	(0.0102)	(0.00793)	(0.00714)	(0.00663)
Observations	52,320	62,825	73,323	83,834
R^2	0.387	0.388	0.387	0.384

sus longer-term behavior. The frequency of the outcomes we investigate is quarterly. As discussed earlier, the HEES provides personalized feedback and energy conservation information. The survey does not provide repeated interaction, as in other HERs such as Opower energy reports. Thus, we are also interested in how customers respond to HEES audit programs in the months or year after the surveys.

Table 4a shows that households did not immediately respond to the non-binding personalized feedback. The average treatment effects increase gradually as time passes.[21] There is no effect after the first three months. The effect after six months is about −3%, and it is approximately −6% nine months later. There is a 7% reduction in electricity consumption one year later compared to households that have not yet joined the program. One year later the treatment behavior does not attenuate, but instead habitual behavior changes. However, there are diminishing returns as time passes. If we evaluate our conclusions together with the results from Du et al. (2014), the contrasting results from Allcott and Rogers (2014) are not surprising. Du et al. (2014) compare the probability of participating in future efficiency programs at six and 12 months and find results of about −4% versus −6%. Households may engage in other energy efficiency programs and are also more likely to reduce their electricity consumption.

Electricity prices are not salient (Shin, 1985; Sallee, 2014). The non-saliency makes incentives ineffective for consumers to change their electricity consumption behavior. Utility consumers in the United States only think about their electricity consumption nine minutes per year, and their attention and interaction increase

---

21 Allcott and Rogers' (2014) Opower study suggests that there is an immediate response to the initial reports. In other words, consumers adjust their behaviors that are feasible in the short term, such as turning off lights and unplugging unused electronics; however, soon there is a "backslide" to pre-intervention consumption levels (Allcott & Rogers, 2014). This suggests that HEES feedback facilitates learning and habitual change.

Table 4b   Over time Kernel propensity-score-matching DID estimations: Treatment effect of participating the January 2009 survey compared to waiting until January 2010. ***Mail-in survey.*** Dependent Variable: Log Consumption.

	(1)	(2)	(3)	(4)
Variables	**Ln(kWh)**	**Ln(kWh)**	**Ln(kWh)**	**Ln(kWh)**
Diff-in-Diff – M	−0.000240	−0.0158**	−0.0458***	−0.0438***
	(0.0107)	(0.00746)	(0.00648)	(0.00591)
Observations	43,451	52,343	60,780	69,514
R^2	0.362	0.362	0.365	0.360

Table 4c   Over time Kernel propensity-score-matching DID estimations: Treatment effect of participating the January 2009 survey compared to waiting until January 2010. ***Online Survey.*** Dependent Variable: Log Consumption.

	(1)	(2)	(3)	(4)
Variables	**Ln(kWh)**	**Ln(kWh)**	**Ln(kWh)**	**Ln(kWh)**
Diff-in-Diff – O	−0.0266	−0.0670***	−0.0964***	−0.110***
	(0.0268)	(0.0209)	(0.0190)	(0.0175)
Observations	9,372	11,278	13,205	15,131
R^2	0.369	0.375	0.372	0.371

Note: For Tables 4a, 4b and 4c, standard errors are in parentheses. Estimations are adjusted for seasonality using seasonal dummies. Time in quarters, from survey participation. Model 1 – effect on 1st quarter, Model 2 – 2 quarters, Model 3 – 3 quarters, and Model 4 – after the entire period (year). *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

when they receive a high bill (Accenture, 2012). The results in Table 4a suggest that receiving higher bills or other intrinsic motivations could cause consumers to pay more attention and curb their incentives by participating in energy efficiency surveys, which could lead to more effective habitual behavioral changes than those of consumers who have not yet participated in the survey.

Tables 4b and 4c present the differential performances of mail-in and online survey participants over time. In Table 4b, which shows the effect for mail-in survey participants, there are immediate reactions to the surveys after the first three months. In the following quarters, there are about 2%, 5%, and 4% reductions in electricity consumption. The disaggregated reactions decrease at a decreasing level. As shown in Table 4c, online survey participants reduced their consumption by about 7%, 10%, and 11% over time. It could be even more interesting if longer-range consumption data were available.

Table 5a Kernel propensity-score-matching QDID estimation of the *all three survey delivery mechanisms*. Quantile DID regression estimates were estimated for the 0.1, 0.25, 0.5, 0.75, and 0.9 quantiles. Dependent Variable: Log Consumption.

	Quantile DID estimates		
Quantile	QDID – Overall	QDID – Mail	QDID – Online
0.1	−0.0815***	−0.0631***	−0.117***
	(0.0135)	(0.0104)	(0.027)
0.25	−0.0562***	−0.0348***	−0.0922***
	(0.00908)	(0.00769)	(0.0179)
0.5	−0.0487***	−0.0351***	−0.0721***
	(0.00815)	(0.00739)	(0.0174)
0.75	−0.0368***	−0.0193**	−0.0706***
	(0.00889)	(0.00853)	(0.0181)
0.9	−0.0298***	−0.0013	−0.0754***
	(0.0111)	(0.0121)	(0.0248)

Note: Standard errors are in parentheses. Estimations are adjusted for seasonality using seasonal dummies. We also tested the equality of coefficients, and the differences between coefficients are statistically significant.
*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Thus far we have discussed the average treatment effect of program participation and have described the average effect of a survey on the typical utility customer. However, because the dependent variable has a continuous distribution, averages may not properly reveal the changes in the distributions (Angrist & Pischke, 2009). Despite the significance of the average effect, we must evaluate whether the magnitude of the effect is persistent and constant for different quantiles. This will show how households at different quantiles may react differently to personalized feedback. We employed QDID by using kernel-based propensity-score-matching estimation, which is an informative framework for examining how the quantiles of energy consumption change in response to survey participation.

Tables 5a and 5b display the QDID estimators and the effects of survey participation at both the central (0.25, 0.5, 0.75) and extreme (0.1 and 0.9) quantiles. Table 5a provides results in terms of percentage changes in the outcome, and Table 5b, as supplemental, provides results for absolute changes in response to the survey participation. Our discussion primarily focuses on percentages or proportional changes. The estimates show significant effects of audit participation compared to households that have not yet participated in the survey. The estimated marginal effects of each quantile regression differ; the estimated marginal effect decreases the farther away one is from the lowest quantile (see Figure 3). House-

Table 5b Kernel propensity-score-matching QDID estimation of the *combined survey effect (with mail-in and online version of the surveys)*. Quantile DID regression estimates were estimated for the 0.1, 0.25, 0.5, 0.75, and 0.9 quantiles. Dependent Variable: kWh Consumption.

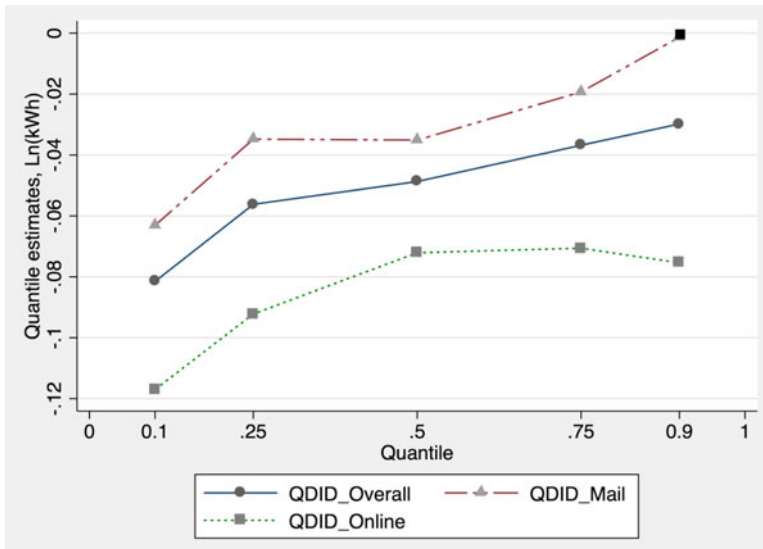|	QDID estimates		
Quantile	QDID – Overall	QDID – Mail	QDID – Online
0.1	−61.92***	−47.13***	−78.85***
	(8.258)	(7.704)	(13.33)
0.25	−43.15***	−45.41***	−69.55***
	(16.87)	(8.922)	(11.56)
0.5	−47.73***	−40.42***	−58.25***
	(9.019)	(9.487)	(13.44)
0.75	−60.07***	−48.86***	−78.42***
	(12.57)	(12.6)	(19.56)
0.9	−52.89**	−17.04	−95.56***
	(20.81)	(21.64)	(29.72)

Note: Standard errors are in parentheses. Estimations are adjusted for seasonality using seasonal dummies.
*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

holds in the lowest quantile save approximately 8% one year after HEES participation, whereas the savings are 3% at the 90th percentile (Column 1, Table 5a). Columns 2 and 3 show the differential performance of the delivery mechanisms. At the extreme quantile of 0.9, there is no evidence of a treatment effect for the mail-in participants. The first three columns of Table 5a also reveal that the lower consumption quantiles saved much more than the upper quantiles among the survey participants versus the comparison group. The changes in the marginal effects for the online audit participants are lower than for the mail-in participants.

The results in Table 5a have additional important implications for policy makers and program designers than simply considering the average effect. Consumers who are in the lowest quantiles are inclined to have more substantial reactions to non-binding energy conservation than consumers in the median or highest quantiles. The critical result of our quantile regressions is showing that the estimated survey effects differ by the level of pre-survey household consumption.[22]

---

22 We computed bootstrapped standard errors for the same regressions to check the robustness and replicability of the results further. Bootstrapped standard errors for matching DID and QDID (not shown) are very similar to the analytical ones we tabulate and yield the same conclusions. We performed 100 bootstrap replications for estimates of the standard errors, which is adequate for normal-approximation confidence intervals (Poi, 2004; Cameron & Trivedi, 2010).

Note: At the 90th percentile, the coefficient for mail-in participants is not statistically significant.

Figure 3 Matching QDID estimates of efficiency program participation effects. Dependent Variable: Log Consumption.

Our final set of results indicate that analyzing distributional impacts rather than just the average effect can provide better understanding about the program effects, and are less limiting in terms of its implications. The pattern of distributional impacts of the energy efficiency program can be a powerful tool to help to assign more targeted and salient interventions in maximizing the program impact and reducing the cost of the implementation of such programs. There is also a discussion in the literature that energy efficiency programs and standards have regressive implications (Fullerton, 2008; Levinson, 2016). Our results show that low-usage households – among the survey participants – save proportionally more energy than do high-use customers. This suggests that once households opt into home energy efficiency programs, such as HEES, there is no evidence of distributionally regressive implications for electricity use here.

5 Concluding remarks and policy implications

Energy efficiency plays a critical role in energy policy debates because meeting our future needs boils down to only two options: increasing supply and decreasing the demand for energy (Gillingham et al., 2006). Due to the high up-front cost of

constructing of large renewable energy facilities, transmission lines and uncertainty in federal or state level support, end-use programs could lessen the pressure by reducing the demand (Considine & Manderson, 2014). Information provision and salience has been documented to affect consumer decisions and decisions to invest in energy-efficient technologies (Newell & Siikamäki, 2014). In our research we examine one of the statewide programs in California (HEES) and determine how well the program has worked in terms of saving energy.

Objective of our study is to provide an alternative measurement approach in evaluating energy efficiency programs. In implementing the method suggested by Sianesi (2004, 2008), we determined the adequate comparison group to correct for the self-selection in non-experimental energy efficiency program evaluations. We then employed a diagnostic test similar to the method suggested by Austin (2009) after matching with estimated kernel-based propensity scores. Combining the two regression estimators, we then described ways to address the systematic differences between the treated and comparison individuals in investigating the effects of residential energy efficiency surveys. Our research is unique in applying the combined methods in evaluating residential energy efficiency programs.

Although the impact was heterogeneous, we provide evidence that the customers who participated in the survey reduced their electricity consumption by about 7%, or about 76 kWh/month on average (see Table 3b). Here, we present a simple calculation of the realized monetary savings for the 2009 HEES participants. If we scale the savings to the entire 2009 HEES participants (January through December) the total reduction in energy consumption would be about 2 million kWh per month, an amount equal to the typical monthly consumption of approximately 3500 households in California. Using a carbon price of $21 per metric ton of carbon dioxide (EPA, 2015; Greenstone et al., 2013), the electricity savings resulted in a monthly estimated reduced emission of 1527 metric tons of carbon dioxide, which is a social cost reduction of about $32,000 per month.[23]

Additionally, we evaluated benefits and costs of the program by using per unit survey cost of the program, as we did not have access to the total cost of the program for 2009, which also included administrative, implementation, measurement, evaluation and other program related costs. So, per unit cost of the mail-in survey was about $12 for SCE's HEES program and let us assume that all the customers used only mail-in surveys, which are more expensive than online surveys.

---

[23] We could not find actual amounts spent for 2009 survey channels. Thus, we rely on the per unit average costs identified for each channel/delivery mechanisms from the "PROCESS EVALUATION FOR THE 2004–2005 STATEWIDE HOME ENERGY EFFICIENCY SURVEY PROGRAM (HEES)" by Opinion Dynamics (2007). The report shows that, for instance, for SCE per unit cost of online surveys are around $11, and mail-in surveys are about $12. For Pacific Gas and Electric (PG&E), they are $20 and $25.

Aggregate (per unit) cost would have been about $324,000, then. On the other hand, using 2011 average residential electricity rates of California (EIA, 2011), which was about 15 cents/kWh, total reduction in monthly bills would have been about $308,000/month. Thus, participating in the voluntary energy efficiency program we examined would save about $12 a month per customer. The net reductions could be higher if customers moved from higher tier rates to lower tiers as a result of reductions in their energy usage. However, since the sample is not randomly selected, it is not possible to infer any tier movement savings to the IOUs' entire high-usage customer base.

The effects of the two versions of the survey both show significant energy-saving effects of 11% and 4% for online and mail-in participants. An implication is that how the program is delivered matters as much as having a program. Du et al. (2014) report similar findings. In addition, our results suggest program effects that become significant and increase in magnitude gradually over time but at a decreasing increment.

Electricity prices are not salient, which already creates a weak incentive to change behavior and routines. To produce more persistent effects customers should be reminded of the intervention because the effects decay. Harding & Hsiaw (2014) suggest that some households may actually view energy efficiency surveys as a commitment device. It is therefore necessary to have additional interactions with households. Because the persistence of treatment also has a spillover effect for the year after the intervention and leads customers to other energy efficiency programs, an assessment of cost-effectiveness should include them too (Allcott & Rogers, 2014).

In addition, because of the heterogeneity in pre-treatment energy consumption, we examined the QDID estimator. Our results suggest that as the quantiles of the distributions increase the effect of the program on electricity consumption decreases in terms of proportionate usage (see Figure 3). Households at the lower quantiles save proportionally more electricity than do customers at higher quantiles.[24] Better customer targeting based on the usage distributions would create significant savings, which would also improve the efficiency of the programs and may also address equity concerns.[25] This suggests that once households opt into HEES, we have no evidence of the program burdening low-use and low-income more than the households in the higher quantile of electricity consumption. Our results imply that program designers can better target low-use and low-income households,

---

[24] However, concerning kWh reduction, smaller percentage change reductions by households in the higher quantiles can achieve more savings of kWh (see Table 5b). This could be due to the kWh difference between low-use and high-use program participants (see Figure 1).

[25] In our program survey sample, only about 13% of participants have incomes below $50,000.

because they are more likely to benefit from such programs through savings. Overall, program participants on average use more electricity than the non-participants (see Figure 1).

We show that understanding of distributional effects can be crucial during the decision of implementing energy efficiency programs and extracting savings cost effectively. Furthermore, better-targeted information can elicit biased beliefs, present bias, inattention, and other decision biases (Allcott, 2016; Allcott & Taubinsky, 2015; Keefer & Rustamov, 2017). Allcott et al. (2015) also suggest that if restricting eligibility is not institutionally feasible, targeted marketing at high-response groups would also generate savings and can enhance policy cost-effectiveness. To a certain degree targeted programs may address the market failures, which are also caused by behavioral biases.

Appendix

Table A1 Variable description and descriptive statistics.

Variable name	Description/range of values	Mean	St.D.
kWh usage	Monthly electricity consumption in kWh data from 2008 January to 2009 December for 4173 customers from all income groups (99568 observations).	1271	639
Bill amount	Monthly Billing information of the randomly selected customers.	236	153
Mail-in	Monthly Billing information for households for mail-in participants.	247.6	144
Online	Monthly Billing information for households of online survey participants.	179.7	184
Income	Income is estimated by the IOU. It is a categorical variable with thirteen different income brackets. Minimum is $1–$15,000, maximum is $250,000. In this study, incomes greater than $100,000 merged in single category.	7.23	
Household size	Number of people in the household. Values range from 1 to 8.	3.59	1.88
Rooms	Number of room in the house. Values range from 1 to 8.	7.9	2.9
Square foot	Size of the house, in square foot.	2480	1076
CDD	Cooling degree days (at 72F).	77	111.3

Table A2 Descriptive Statistics within survey participants.

Variables	Survey types – household characteristics	
	Mail-in	**Online**
kWh – pre-survey mean	1355.7	1001.9
Treatment	1468.6	946.9
Comparison	1336.6	1068.2
Household size – mean	3.67	3.16
House ownership	96%	87%
Income (1–13) – mean	7.24	7.13
Weather (CDD) – mean	76.51	79.56
# Rooms – mean	8.01	7.4
House type		
1-2-3 Story house vs. apartment/condo	98.25%	92.15%

Table A3 Descriptive Statistics for Non-participant sample (vs. Participants)

Name	Description	Random accounts			Survey participants		
		Mean	S.D.	%	Mean	S.D.	%
Household size	Number of people in the household. Values range from 1 to 8.	2.78	1.94		3.6	1.89	
Rent status	% of the observations. For random sample about 20% of the observations' ownership status was not determined.			6.1%			1.1%
Income	The IOU estimated income. It was a categorical variable with 13 income brackets.	5.8			7.2		
Climate zone	% of the sample shared the same climate zones.			88%			91%

Table A4   Sample of questions from the HEES survey.

Survey questions	Yes (%)
Portion of the year home is occupied?	99.73
Percentage of customer owns house or condo?	78
Are you considering remodeling home?	14
Do you have a whole house fan?	80
Do you have a dishwasher?	82
Do you use energy-saving showerheads?	87
Do you let the water faucet run when you brush your teeth?	60
Do you let the water faucet run when you wash your face and hands?	81
Do you let the water faucet run when you clean dishes?	81
Do you let the water faucet run when you prepare food?	93
Do you have a pool?	52

Note: There are more than 130 questions in the survey. Since the objective is not questions, but the subsequent behavior of survey participation, we just provide some sample questions and their statistics.

References

Abrahamse, Wokje, Steg, Linda, Vlek, Charles & Rothengatter, Talib (2005). A Review of Intervention Studies Aimed at Household Energy Conservation. *Journal of Environmental Psychology*, *25*(3), 273–291.

Accenture (2012). Actionable Insights for the New Energy Consumer. *Accenture End-Consumer Observatory* 2012.

Akerlof, George A. (1978). The Economics of 'Tagging' As Applied to the Optimal Income Tax, Welfare Programs, and Manpower Planning. *American Economic Review*, *68*(1), 8–19.

Allcott, Hunt (2011). Social Norms and Energy Conservation. *Journal of Public Economics*, *95*(9–10), 1082–1095.

Allcott, Hunt (2016). Paternalism and Energy Efficiency: An Overview. *Annual Review of Economics*, *8*, 145–176.

Allcott, Hunt & Kessler, Judd (2018). The Welfare Effects of Nudges: A Case Study of Energy Use Social Comparisons. *American Economic Journal: Applied Economics*, forthcoming.

Allcott, Hunt, Knittel, Christopher & Taubinsky, Dmitry (2015). Tagging and Targeting of Energy Efficiency Subsidies. *American Economic Review: Papers and Proceedings*, *105*(5), 187–191.

Allcott, Hunt & Mullainathan, Sendhil (2010). Behavioral Science and Energy Policy. *Science*, *327*(5970), 1204–1205.

Allcott, Hunt & Rogers, Todd (2014). The Short-Run and Long-Run Effects of Behavioral Interventions: Experimental Evidence from Energy Conservation. *American Economic Review*, *104*(10), 3003–3037.

Allcott, Hunt & Taubinsky, Dmitry (2015). Evaluating Behaviorally-Motivated Policies: Experimental Evidence from the Light Bulb Market. *American Economic Review*, *105*(8), 2501–2538.

Angrist, Joshua D. & Pischke, Jörn-Steffen (2009). *Mostly Harmless Econometrics*. Princeton, NJ: Princeton University Press.

Athey, Susan & Imbens, Guido W. (2006). Identification and Inference in Nonlinear Difference-in-Difference Models. *Econometrica*, *74*(2), 431–497.

Austin, Peter C. (2009). Balance Diagnostics for Comparing the Distribution of Baseline Covariates Between Treatment Groups in Propensity-Score Matched Samples. *Statistics in Medicine*, *28*, 3083–3107.

Baser, Onur (2006). Too Much Ado about Propensity Score Models? Comparing Methods of Propensity Score Matching. *Value in Health*, *9*(6), 377–385.

Black, Nick (1996). Why We Need Observational Studies to Evaluate the Effectiveness of Health Care. *BMJ*, *312*(7040), 1215–1218. May 11.

Black, Dan A. & Smith, Jeffrey A. (2004). How Robust is the Evidence on the Effects of College Quality. *Journal of Econometrics*, *121*, 99–124.

Blundell, Richard & Costa Dias, Monica (2009). Alternative Approaches to Evaluation in Empirical Microeconomics. *Journal of Human Resources*, *44*(3), 565–640.

Borenstein, Severin (2017). The Private Net Benefits of Residential Solar PV: The Role of Electricity Tariffs, Tax Incentives and Rebates. *Journal of the Association of Environmental and Resource Economists*, *4*(S1, Part 2), S85–S122.

Brennan, Timothy J. (2010). Decoupling in Electric Utilities. *Journal of Regulatory Economics*, *38*(1), 49–69.

Caliendo, Marco & Kopeinig, Sabine (2005). Some Practical Guidance for the Implementation of Propensity Score Matching. IZA DP No. 1588.

Cameron, Colin A. & Trivedi, Pravin K. (2010). *Microeconometrics Using Stat. Rev. ed.* College Station, TX: Stata Press.

Charness, Gary & Gneezy, Uri (2009). Incentives to Exercise. *Econometrica*, *77*(3), 909–931.

Chen, Caroline, Rustamov, Galib, Hirsch, Kelly, Lau, Kenneth, Buendia, Jose & Ayuyao, Eugene (2015). 10 (Electricity)-10(Gas)-10(Water) Plus Multi-family Competition and Energy Star Portfolio Manager Benchmarking Pilot Program Design. In *ACEEE Summer Study on Energy Efficiency in Industry*.

Considine, Timothy & Manderson, Edward (2014). The Role of Energy Conservation and Natural Gas Prices in the Costs of Achieving California's Renewable Energy Goals. *Energy Economics*, *44*, 291–301.

Davis, Lucas W., Fuchs, Alan & Gertler, Paul (2014). Cash for Coolers: Evaluating a Large-Scale Appliance Replacement Program in Mexico. *American Economic Journal: Economic Policy*, *6*(4), 207–238.

Deaton, Angus (2000). In *The Analysis of Household Surveys: A Microeconomics Approach to Development Policy*. (3rd ed.). Baltimore: The John Hopkins University Press; Published for the World Bank.

Dehejia, Rajeev H. & Wahba, Sadek (2002). Propensity Score-Matching Methods for Nonexperimental Causal Studies. *The Review of Economics and Statistics*, *84*(1), 151–161.

Delmas, Magali A., Fischlein, Miriam & Asensio, Omar I. (2013). Information Strategies and Energy Conservation Behavior: A Meta-Analysis of Experimental Studies from 1975 to 2012. *Energy Policy*, *61*, 729–739.

Du, Yingjuan, Hanna, Dave, Shelton, Jean & Buege, Amy (2014). What Behaviors Do Behavior Programs Change. In *ACEEE Summer Study on Energy Efficiency in Buildings*.

Dubin, Jeffrey A., Miedema, Allen K. & Chandran, Ram V. (1986). Price Effects of Energy-Efficient Technologies: A Study of Residential Demand for Heating and Cooling. *RAND Journal of Economics*, *17*(3), 310–325.

ECONorthwest (2009). Process Evaluation of the SCE 2006-08 Home Energy Efficiency Survey (HEES) Program: Final Report. Study ID: SCE0275.01.

Fowlie, Meredith, Greenstone, Michael & Wolfram, Catherine (2015). Are the Non-Monetary Costs of Energy Efficiency Investments Large? Understanding Low Take-up of a Free Energy Efficiency Program. *American Economic Review: Papers and Proceedings*, *105*(5), 201–204.

Frölich, Markus & Melly, Blaise (2010). Estimation of Quantile Treatment effects with Stata. *The Stata Journal*, *10*(3), 423–457.

Fu, Alex Z., Dow, William H. & Liu, Gordon G. (2007). Propensity Score and Difference-in-Difference Methods: A Study of Second-Generation Antidepressant Use in Patients with Bipolar Disorder. *Health Service and Outcomes Research Methodology*, *7*, 23–38.

Fullerton, Don (2008). *Distributional Effects of Environmental and Energy Policy: An Introduction*. NBER Working Paper 14241.

Gayer, Ted & Viscusi, W. Kip (2013). Overriding Consumer Preferences with Energy Regulations. *Journal of Regulatory Economics*, *43*(3), 248–264.

Gerarden, Todd D., Newell, Richard G. & Stavins, Robert N. (2017). Assessing the Energy-Efficiency Gap. *Journal of Economic Literature*, *55*(4), 1486–1525.

Gillingham, Kenneth, Newell, Richard & Palmer, Karen (2006). Energy Efficiency Policies: A Retrospective Examination. *Annual Review of Environment and Resources*, *31*, 161–192.

Gillingham, Kenneth, Newell, Richard & Palmer, Karen (2009). Energy Efficiency Economics and Policy. *Annual Review of Resource Economics*, *1*, 597–620.

Greenstone, Michael, Kopits, Elizabeth & Wolverton, Ann (2013). Developing a Social Cost of Carbon for U.S. Regulatory Analysis: A Methodology and Interpretation. *Review of Environmental Economics and Policy*, *7*(1), 23–46.

Hamermesh, Daniel S. & Trejo, Stephen S. (2000). The Demand for Hours of Labor: Direct Evidence from California. *The Review of Economics and Statistics*, *82*(1), 38–47.

Hansen, Ben B. (2004). Full Matching in an Observational Study of Coaching for the SAT. *Journal of the American Statistical Association*, *99*(467), 609–618.

Harding, M. & Hsiaw, A. (2014). Goal Setting and Energy Conservation. *Journal of Economic Behavior and Organization*, *107*, 209–227.

Heckman, James, Ichimura, Hidehiko, Smith, Jeffrey & Todd, Petra (1998). Characterizing Selection Bias Using Experimental Data. *Econometrica*, *66*(5), 1017–1098.

Heckman, James, LaLonde, Robert & Smith, Jeffrey (1999). The Economics and Econometrics of Active Labor Market Programs. In Orley Ashenfelter & David Card (Eds.), *Handbook of Labor Economics Vol. III* (pp. 1865–2097). Amsterdam: Elsevier.

Imbens, Guido W. & Rubin, Donald B. (2015). *Causal Inference for Statistics, Social, and Biomedical Sciences: An Introduction*. New York, NY: Cambridge University Press.

Ingle, Aaron, Moezzi, Mithra M., Lutzenhiser, Loren, Hathaway, Zac, Lutzenhiser, Susan, Joe, Van Clock, Peters, Jane S., Smith, Rebecca, Heslam, David & Diamond, Richard (2012). "Behavioral Perspectives on Home Energy Audits: The Role of Auditors, Labels, Reports, and Audit Tools on Homeowner Decision-Making". *Lawrence Berkeley National Laboratory*, LBNL-5715E.

Itron, Inc. (2013). 2010–2012 CPUC HEES Impact Evaluation: Final Report. Prepared for California Public Utilities Commission.

Keefer, Quinn & Rustamov, Galib (2017). Limited Attention in Residential Energy Markets: A Regression Discontinuity Approach. *Empirical Economics*, https://doi.org/10.1007/s 00181-017-1314-6.

Kennedy, Peter (2003). *A Guide to Econometrics*. (5th ed.). Cambridge, MA: The MIT Press.

Lechner, Michael (2002). Some Practical Issues in the evaluation of heterogeneous Labour Market Programmes by Matching Methods. *Journal of the Royal Statistical Society*, *165*, 59–82.

Levinson, Arik (2016). *Energy Efficiency Standards Are More Regressive Than Energy Taxes: Theory and Evidence*. NBER Working Paper 22956.

List, John A., Millimet, Daniel L., Fredriksson, Per G. & McHone, Warren W. (2003). Effects of Environmental Regulations on Manufacturing Plant Births: Evidence from a Propensity Score Matching Estimator. *The Review of Economics and Statistics*, *85*(4), 944–952.

Meyer, Bruce D. (1995). Natural and Quasi-Experiments in Economics. *Journal of Business and Economic Statistics*, *13*(2), 151–161.

Meyer, Bruce D., Viscusi, W. Kip & Durbin, David L. (1995). 'Workers' Compensation and Injury Duration: Evidence from a Natural Experiment. *The American Economic Review*, *85*(3), 322–340.

Nadel, Steve & Keating, Kenneth (1991). Engineering Estimates vs. Impact Evaluation Results: How Do They Compare and Why? Research Report U915. Washington, D.C.: American Council for an Energy-Efficient Economy. http://www.aceee.org/research-re port/u915.

Newell, Richard & Siikamäki, Juah (2014). Nudging Energy Efficiency Behavior: The Role of Information Labels. *Journal of the Association of Environmental and Resource Economists*, *1*(4), 33–38.

Poi, Brian P. (2004). From the Help Desk: Some Bootstrapping Techniques. *Stata Journal*, *4*, 321–328.

Rosenbaum, Paul R. & Rubin, Donald B. (1983). The Central Role of the Propensity Score in Observational Studies for Causal Effects. *Biometrica*, *70*(1), 41–55.

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

Rubin, Donald B. & Thomas, Neal (1992). Affinely Invariant Matching Methods with Ellipsoidal Distributions. *The Annals of Statistics*, *20*(2), 1079–1093.

Sallee, James M. (2014). Rational Inattention and Energy Efficiency. *Journal of Law and Economics*, *57*(3), 781–820.

Shin, Jeong-Shik (1985). Perception of Price When Price Information Is Costly: Evidence from Residential Electricity Demand. *The Review of Economics and Statistics*, *67*(4), 591–598.

Sianesi, Barbara (2004). An Evaluation of the Swedish System of Active Labour Market Programs in the 1990s. *The Review of Economics and Statistics*, *86*(1), 133–155.

Sianesi, Barbara (2008). Differential Affects of Active Labour Market Programs for the Unemployed. *Labour Economics*, *15*, 370–399.

Smith, Herbert L. (1997). Matching with Multiple Controls to Estimate Treatment Effects in Observational Studies. *Sociological Methodology*, *27*, 325–353.

Smith, J. A. & Todd, P. E. (2001). Reconciling Conflicting Evidence on the Performance of Propensity Score Matching Methods. *American Economic Review, Papers and Proceedings*, *91*, 112–118.

The Cadmus Group, Inc. (2017). Evaluation of Southern California Edison 2015 Home Energy Efficiency Survey Programs. Draft Report. California Public Utilities Commission.

Titus, Marvin A. (2007). Detecting Selection Bias, Using Propensity Score Matching, and Estimating Treatment Effects: An Application to the Private Returns to a Master's Degree. *Research in Higher Education*, *48*, 487–521.

U.S. Energy Information Administration (2011). Electric Power Monthly, January 2011.

U.S. Environmental Protection Agency (2015). Greenhouse Gas Equivalencies Calculator. Retrieved from on January 15 from http://www.epa.gov/energy/greenhouse-gas-equivalencies-calculator, last updated on October 23.

Wirl, Franz & Orasch, Wolfgang (1998). Analysis of United States' Utility Conservation Programs. *Review of Industrial Organization*, *13*(4), 467–486.