DISCUSSION PAPER SERIES

IZA

IZA DP No. 9567

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

Thomas J. Kniesner
Galib Rustamov

December 2015

Forschungsinstitut
zur Zukunft der Arbeit
Institute for the Study
of Labor

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

Thomas J. Kniesner
Claremont Graduate University,
Syracuse University (Emeritus) and IZA

Galib Rustamov
Claremont Graduate University

Discussion Paper No. 9567
December 2015

IZA

P.O. Box 7240
53072 Bonn
Germany

Phone: +49-228-3894-0
Fax: +49-228-3894-180
E-mail: iza@iza.org

Any opinions expressed here are those of the author(s) and not those of IZA. Research published in this series may include views on policy, but the institute itself takes no institutional policy positions. The IZA research network is committed to the IZA Guiding Principles of Research Integrity.

The Institute for the Study of Labor (IZA) in Bonn is a local and virtual international research center and a place of communication between science, politics and business. IZA is an independent nonprofit organization supported by Deutsche Post Foundation. The center is associated with the University of Bonn and offers a stimulating research environment through its international network, workshops and conferences, data service, project support, research visits and doctoral program. IZA engages in (i) original and internationally competitive research in all fields of labor economics, (ii) development of policy concepts, and (iii) dissemination of research results and concepts to the interested public.

IZA Discussion Papers often represent preliminary work and are circulated to encourage discussion. Citation of such a paper should account for its provisional character. A revised version may be available directly from the author.

IZA Discussion Paper No. 9567
December 2015

ABSTRACT

Differential and Distributional Effects of Energy Efficiency Surveys: Evidence from Electricity Consumption*

Our research investigates the magnitude of the effect of residential energy efficiency audit programs on later household electricity consumption. These programs are designed to increase awareness of household energy consumption with personalized feedback that will eventually lead to behavioral changes. In this type of survey, there is only a one-time interaction between households, which participate voluntarily, and the surveyors. The objective of this study is to determine whether and to what extent such surveys lead to behavioral changes. We argue that the perceived complexity of the survey feedback will determine whether the subsequent behavior is sustainable. Then we analyze how persistent the intervention is over time and whether the effects decay or intensify. However, the main evaluation problem involving these surveys is self-selection bias. To correct for this bias, we propose two non-parametric estimators by using a kernel-based propensity score matching approach. In the first method, we use "difference-indifferences" (DID) estimations. The 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. The evidence suggest that the customers who participated in the survey reduced their electricity consumption by 6.7%, compared with customers who had not yet participated in the survey. In addition, as the quantiles of the distribution increase, the effect of the program decreases.

JEL Classification: C31, D03, D12, L94, Q41

Keywords: electricity consumption, information salience, selection bias, propensity-score matching, treatment effects, multiple treatments

Corresponding author:

Thomas J. Kniesner
Claremont Graduate University
Harper East 216
Claremont, CA 91711
USA
E-mail: thomas.kniesner@cgu.edu

* We thank Hal Nelson, C. Monica Capra, Joshua Tasoff, Quinn Keefer, and Shahana Samiullah for their helpful comments.

1. Introduction

Home energy audits have been offered in the United States since at least the 1970s, and the use of these audits has expanded with the use of stimulus funds in recent years (Ingle et al., 2012). In California, home energy efficiency survey (HEES) programs are implemented statewide by the public utilities. The objectives of these programs 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 money and energy. Similarly to other types of surveys, home energy audits provide a rich source of data on energy consumption. These 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, it is expected that personalized feedback will lead to the desired behavioral change. "Being surveyed can change subsequent behavior and related parameter estimates" (Zwane et al., 2011).

As discussed in the earlier studies, individuals may behave inefficiently because of the unclear relationship between price and behavior in electricity consumption. Thus, home energy efficiency audits are expected to close the information gap (with personalized feedback) by serving as a reminder. Therefore, personalized feedback may decrease information asymmetry and result in more efficient behavior and actions. The survey data that we examine in this paper are non-experimental; thus, the samples in the data were not randomly selected from a controlled environment. "The self-selection is likely to be associated with other important differences that exist between participant and non-participant households that could help explain the participation choice and associated energy efficiency program participation choices of these households" (Du et al., 2014). One would expect more marked changes in the energy efficiency behavior of the participants who self-select.

In addition to the self-selection issue, there is also an incentive issue regarding the way energy audit programs have been implemented. Utility companies have been one of the major implementers of home energy audit programs. Under regulatory practice, utilities have an incentive to invest in conservation measures, but they limit actual conservation through the improper design of a program (Wirl and Orasch, 1998). In addition, as part of the American Recovery and Reinvestment Act of 2009, the U.S. Department of Energy (DOE), acting through its Office of Energy Efficiency and Renewable Energy, increased its support of residential energy efficiency technologies and programs (Ingle et al., 2012). All investor-owned electric and gas utilities in California engage in decoupling. "[D]ecoupling, which separates electricity retailers' profits from quantities sold, is one mechanism that could encourage firms to nudge consumers toward reducing energy usage" (Brennan, 2010; Allcott and Mullainathan, 2010). Specifically, decoupling does not provide an affirmative incentive for utilities to encourage conservation; it simply removes a disincentive not to conserve. Because utilities have been under regulatory pressure to take such measures and they are (generally) the energy suppliers, 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 base. Therefore, although we may encounter some successes, overall, the ways in which home energy audits have been designed and executed have often been ineffective, such that there are lower response rates to such surveys. Moreover, utilities have not used scientific methods to implement their energy efficiency programs; thus, experimental or quasi-experimental approaches have not been part of the incentives. Therefore, self-selected studies did not particularly 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. Recently in California, the CPUC made it mandatory for all statewide IOUs to implement behavior-based programs. An example of such a behavior-based program is the implementation of social comparisons by OPOWER (see, e.g., Allcott 2011).

"Much of the empirical microeconomic literature [in development] uses econometric and statistical methodology to overcome the non-experimental nature of data" (Deaton, 2000). Thus, because of the inherent selection bias, we began the analysis by employing the empirical technique suggested by Sianesi (2004), who examines the effectiveness of unemployment programs in Sweden. Sianesi (2004) suggests selecting future program participants for matching estimations. We apply the method in a different market setting, namely, residential energy efficiency audits. The DID estimator provides evidence that participation in the survey leads to 6.7% less electricity consumption by survey participants than by customers who did not participate. In addition, the effect is persistent over time, at least for the entire year after the survey. Furthermore, we employ quantile regression to detect the effects of home energy surveys on the distributions, not only on individual households. A theoretical discussion adds a slight extension to the model developed by Frondel and Vance (2012).

Finally, utilities use various delivery mechanisms to implement home energy audit programs -- through mail, online, telephone and in-home (on-site) audits. In this paper, we investigate the differential performance of the mail-in compared with the online versions of the home energy surveys in addition to the general impact of the surveys. Because of fewer observations, we excluded the on-site and telephone surveys.

2. Theoretical Framework

Frondel and Vance (2012) provided the conceptual foundations for this section. They investigated how consumers respond to home energy audits, especially "the role of information in influencing decisions about retrofitting" and whether such programs could trigger energy efficient renovations among the participants. Instead, in this study, we infer a continuous or habitual behavioral change in electricity consumption.

As indicated by Frondel and Vance (2012), in the first step of the decision model, customer i decides whether to take the survey. Because customers have imperfect information regarding their own energy consumption behavior prior to the survey, there is an expected benefit, $E(B_i)$, from acquiring personalized feedback from the survey, for example, whether a customer is satisfied with the service, i.e., customer realized inefficient own behavior. Additionally, there is a cost associated with acquiring the information, C_i. This cost may be the time that it takes customers to answer more than 100 questions about their houses and consumption behavior. Therefore, a household decides to participate in the survey if $E(B_i) > C_i > 0$.

At the second stage, the customer decides whether to take action. Prior to taking action to implement the feedback, customer i forms the expectation $E(V_i)$ on the basis of the individual present value V_i, including both the financial and behavioral costs that result from taking action to change the behavior.[1] When $V_i > 0$, the customer decides to respond positively to the feedback.

Finally, in the third stage of this model, we investigate whether customers are persistent in their actions. The primary objective is to present evidence that the behavioral response to the feedback is determined by how a household perceives the information. We define the feedback, ϕ, from the survey as the sum of both the affordable (relevant, cheap and easy to apply tips) information, A, and the unaffordable (non-relevant, expensive and complex tips) information, N. Therefore, the actual valuation of the feedback should be $\phi = A + N$. However, the literature suggests that if the information is complex, it becomes difficult to change the behavior. Then, the perceived valuation of the feedback is $\phi' = \delta * \phi$, where $\delta \in [1, \infty)$ is a complexity parameter.[2] $\delta = 1$ denotes that people do not perceive the personalized feedback to be any more difficult to employ than it is. We assume that public utilities make the survey feedback and information actionable. $\delta > 1$ denotes that people perceive the information to be more difficult to employ than it is. Thus, although consumers may have an initial motivation to change their behavior, the perceived complexity of the survey feedback will determine whether the subsequent behavior is sustainable. As δ moves away from 1, the effects of

[1] V_i depends on i's time preference rate ρ_i, the vector of customer demographic characteristics x_i and uncertain revenues due to energy conservation behaviors (in period t): $V_i = E(V_i) + \varepsilon_i$. For

[2] In contrast to the inattention parameter suggested by DellaVigna (2009) and Sexton (2015), we assume here that the complexity parameter is greater than or equal to 1. The inattention parameter suggests that consumers overweigh the visible component of the (part of price) information (Sexton, 2015).

participation (or initial ambition) will decay more rapidly over time. Eventually, this decay creates inertia in consumers regarding adoption of energy efficient behavior or participation in any future (behavioral) energy efficiency programs.

3. Data

In this section, we discuss the data that an IOU in California provided on a confidential basis. The data are for more than 4,200 customers who participated in the HEES in January of 2009 and 2010. We eliminated any households that had less than 12 months of consumption data during the period, thus leaving 4,173 households. The way that the survey is structured by the program designers suggest that the samples suffer from self-selection bias. To address the research question meaningfully, we chose the January 2009 survey participants as the treatment group and the future survey participants, January 2010, as the comparison group. The comparison group comprises customers who did not participate in January 2009 and have not yet participated in the survey (Sianesi, 2004). The summary statistics for the sample are presented in Table 1. The data set used in here is the result of combining three main sources that reflect monthly energy consumption, namely, billing, dwelling demographics, and the survey (HEES). The billing and demographics components of the variables are the same for randomly selected customers and were explained in the previous paper. The survey data cover the years 2008 and 2009 for both the 2009 and 2010 survey participants. The weather information was collected by using the monthly Cooling Degree-Days (CDD) data over the billing period from 2008-2009 and was merged with the main dataset. Because California has warmer weather than the national average, a 72F indoor baseline temperature was used instead of the nationally defined baseline of 65F.

The Home Energy Efficiency Survey (HEES) program is a resource-acquisition program that provides residential customers with an energy analysis 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 home and then offers a specific list of tips based on their responses. The recommendations include both 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 execute the recommendations. It is important to determine whether the design of the HEES report is successfully imparting useful knowledge and referring participants to helpful resources, and whether this coordination effort is motivating participants to adopt more energy- and water-efficient behaviors. We only focus on mail-in and online survey participant data. These two methods are commonly compared with other methods. Furthermore, telephone and in-home surveys have become less frequently implemented by utilities. In-home data are costly for utilities to collect, although the largest savings are observed as a result of this type of intervention. In addition, because of fewer observations, we did not account for the in-home and telephone participants.

The online survey version is offered in 2 lengths: a standard length, “Energy 15” – long, which is intended to take 15-30 minutes, and a shorter length, “Energy 5” – short, which is intended to take 5 minutes. We combined these two parts of the online version under a single variable. The mail-in surveys have the same questions as the “Energy 15” survey; the only difference is that the customer receives feedback one week later.[3] The Appendix lists the selected survey variables and descriptive statistics.

4. Method

We start the analysis by measuring the impact of the overall HEES audit program participation. Because the audit program uses online-based and mailing delivery mechanisms (or formats) to reach customers, we continue the analysis by evaluating the impact of each format separately on post-audit energy consumption behavior and the magnitudes of each format. In the present study, the treatment group is the January 2009 program participants, and the comparison group is the January 2010 program participants. Because customers opt-in to the programs, there is a selection bias. To address this bias, we first identify the valid comparison group. We chose the January 2010 program participants (future survey participants) as the comparison group. In this scenario, the classical treatment and control distinction clearly holds (Sianesi, 2004). This framework will determine the proper and valid matching estimations. This approach is more reliable (see Sianesi, 2004; 2008) than matching individuals who have never participated in home energy audits (see Du et al., 2014). The next two sections explain the challenges of selection bias and provide the estimation techniques to mitigate the potential self-selection bias.

4.1. Selection Bias in Energy Efficiency Surveys

The objective of this paper – to determine whether a household survey and the method of the survey change the later behavior of participants – is to study the role of customized feedback on the customer's energy consumption behavior and attention level.

---

[3] The program implementers also provide these services in different languages, through Chinese, English, Korean, Spanish, and Vietnamese.

Ideally, to generate valid conclusions and understand the population, it is important to conduct the program with randomized social experiments. Randomized experiments create for independence between the treatment and consumer characteristics, including both observable and unobservable characteristics. Thus, non-randomized observational data can be misleading because of selection bias – decisions made by the households to participate in the energy efficiency survey. The main concerns are the unmeasured factors, such as motivation to take action. These concerns may affect the decision to participate in the survey and may also affect post-intervention performance. Additionally, a customer who has requested an audit may be from the type of household that is taking other unobserved actions to conserve energy (Allcott and Mullainathan, 2010). This confounding difference between the participants and the non-participants shows the difficulty of controlling for these differences when estimating the causal effects of these programs. "The main problem here is that often the researcher wishes to draw conclusions about the wider population, not just the subpopulation from which the data is taken" (Kennedy, 2003). However, because of ethical problems, the large costs of implementing randomizations, and problems with external validity (Fu, Dow, and Liu, 2007; Black, 1996), many studies use observational data instead of implementing a randomized experiment.

Similarly to many other energy efficiency survey programs, the HEES audit program that is used in this research is also evaluated by a non-experimental approach where the customer chooses to participate in the survey instead of being randomly assigned by the program designer. Therefore, the data suffer from selection bias, which makes it difficult to know what the response will be if the program is implemented on a mandatory basis or through some added participation incentive payment. However, if the underlying question is simply "How do *voluntary* participants in these programs respond?" then it seems that there is no selection bias**.** Because the subject of the study focuses on the former question of mandatory implementation, this study also presents solutions to the problem of selection bias. To provide a proper estimate of the treatment effect with observational data, we consider the sample selection phenomenon. Then, we begin the analysis by employing the method that is suggested by Sianesi (2004), where the comparison group consists of customers who were not yet participating in the survey but participated later.[4]

As discussed above, because of the nature of social programs, such as residential energy efficiency surveys, a group that never participated in the survey cannot simply be chosen. Although many studies suggest alternative methods for evaluating social programs where the comparison group was never treated, identifying both the appropriate sample in the nontreated population and the estimator was the main objective of these studies.[5] The most common solution to the selection problem in social program evaluations is the matching approach. The idea is to identify the non-treated individuals in the comparison group who are similar to the individuals in the treatment group in their pre-treatment characteristics. It is difficult to estimate the effects of program participation if the pre-survey variables between the treatment and comparison groups are dissimilar, even if we employ matching on the propensity score. Therefore, instead of using random utility customers as a comparison group and matching them with the treatment group

---

[4] The details are explained in the next sections.

[5] See Heckman (1980), Heckman, Ichimura, Smith and Todd (1998), Heckman, Ichimura and Todd (1997, 1998), and Imbens and Rubin (2015) for a detailed discussion about the matching and addressing the comparison group in social programs.

based on observable pre-survey characteristics, we use customers who joined the program later, in January 2010.

4.2. Evaluation Approach

"Using the mean outcome of untreated individuals $E[Y_0|T = 0]$[6] is in non-experimental studies usually not a good idea because it is most likely that components which determine the treatment decision also determine the outcome variable of interest" (Caliendo and Kopeinig, 2005). This observation suggests that it is likely that even if the best possible candidate for the comparison group is chosen, the consumption levels will still be different even if consumers do not participate in the surveys because of the unobserved counterfactual. We start the analysis by estimating the effect without matching for comparison with the other models. Then, we estimate the effect by employing 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):

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

In this study, because the comparison group is selected 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,

[6] T = 1 if individual *i* participates in the survey in January 2009 (treatment).

2004, 2008; Caliendo and Kopeinig, 2005)[7], which suggests that the probability distributions of the two groups are very similar to each other. Another requirement for the matching methods procedure is the CS or overlap condition:

$$0 < Pr(T = 1|\, X) < 1 \tag{2}$$

"This condition guarantees that persons with the same *X* values have a positive probability of being both participants and non-participants" (Heckman, LaLonde, and Smith, 1999). This condition suggests that for every customer in the treatment group, there are customers with similar characteristics in the comparison group. Heckman, LaLonde and Smith (1999) show that this condition "is central to the validity of matching". Considering these two conditions, the literature suggests that the propensity score is useful to construct the matching estimators. The propensity score is the conditional probability of being treated at time *t* given a vector of observed characteristics to reduce the dimensionality of the matching problem (Rosenbaum and Rubin, 1983). The propensity score estimates the propensity of the customers with the observed characteristics to receive the program -- the energy efficiency survey.[8] Thus, the customers who have the same or similar propensity score values have similar distributions of all of the observable characteristics. Figure 1 shows that the customers in the treatment and comparison groups have similar propensity score distributions. According to Dehejia and Wahba (1999), propensity score matching estimates are more

[7] We use the pre-treatment characteristics of *X* for the CIA.

[8] We use conditional propensity score based on the some of the pre-treatment observable characteristics such as income, weather (CDD), house ownership and type of house customers live in. 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 and Rubin, 2015). The study suggests that it is both difficult and not efficient to employ large number of covariates. Considering both graphical and empirical results the estimated conditional propensity score was appropriate for continuing to calculate the non-parametric estimators.

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 it is 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 and Rubin 1983, 1985; Rubin and 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 individuals in the comparison group to construct the counterfactual outcome" (Caliendo and Kopeinig, 2005). This 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.

Then, we introduce the nonparametric versions of the difference-in-differences (DID) estimation to the later participants as a comparison group that uses 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 by using 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 bias in the sample, a difference-in-differences matching estimator to control for the presence of the unobservable characteristics was used, as referenced in List et al. (2003). Heckman et al. (1998) argue that propensity score DID accounts for the difference between the treatment and the comparison groups, which eliminates the bias. The design for the DID model is as follows. Individual *i* belongs to either the treatment or comparison group, $T_i \in \{0, 1\}$, where $T = 1$ is the treatment group. The period of *i's* consumption behavior is defined as $P_i \in \{0,1\}$. Y_i is the outcome variable – monthly energy consumption in kWh. The interaction term $T_i \cdot P_i$ is an indicator of the treatment. Then, the standard DID model for the realized outcome is the following:

$$Y_i = \alpha + \beta T_i + \gamma P_i + \psi(T_i \cdot P_i) + \theta X + \epsilon_i. \tag{3}$$

The coefficient of the interaction term, ψ, is DID, or the impact of survey participation on later consumption behavior. *X* is a vector of household 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 (see Athey and Imbens, 2006):

$$\psi^{DID} = \mathbb{E}[Y_i|T_i = 1, P_i = 1] - \mathbb{E}[\,[Y_i|T_i = 1, P_i = 0] \\ -\mathbb{E}[Y_i|T_i = 0, P_i = 1] - \mathbb{E}[Y_i|T_i = 0, P_i = 0] \tag{4}$$

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 this study has access to the pre- and post-treatment residential energy consumption data, this approach is suitable for this study.

Another type of non-parametric approach that we test is the quantile DID (QDID) matching method. We continue using kernel-based propensity score matching. The focus for the previous DID method is to produce the average causal effects of program participation. In addition, however, we are also investigating the effect of the programs on the entire distribution. 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 and 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, Viscusi and Durbin, 1995). In this case, the survey has different effects in different quantiles. The QDID estimates are estimated for both the extreme (0.1 and 0.9) and central (0.25, 0.5, 0.75) quantiles. The standard QDID estimator on quantile q can be shown as the following (Athey and Imbens, 2006):

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

where $F_Y^{-1}(q|X)$ is the distribution function for Y at q, which is conditional on the X (the matched observable characteristics or propensity scores). The equation shows the difference between treatment and comparison group before and after the treatment for different quantiles. To our knowledge, our study represents one of the earliest attempts to apply the QDID matching method to residential energy efficiency program evaluations.

We primarily focus on the changes in kWh usage. We also use the natural logarithmic transformation, Ln(kWh), where the interpretation of the effect is in terms of percentage changes. To identify the durability of the intervention, we also capture both short- (quarterly) and long-term (year) effects of energy efficiency survey participation. To examine the validity and to verify the results, we calculated the bootstrapped confidence intervals (Lechner, 2002; Black and Smith, 2003; Sianesi, 2004). These methods can help to improve the validity of the analysis and mitigate the potential bias of the estimation.

5. Results

Participation in the energy audit program is voluntary. If non-participants were chosen as a comparison group, systematic differences would be apparent between the participant and non-participant groups because of unobservable motivation and observable household characteristics.[9] Thus, the analysis initially focuses on identifying and justifying the valid comparison group, and then it continues with the regression estimation. The objective is to prevent an inflated estimate of the potential audit program's impact. The interest in calculating the propensity score and matching (method) "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 start with kernel propensity score matching DID, which produces the average treatment effects. In addition, to investigate the impact of an audit on the entire distribution, we employ the kernel propensity score matching QDID. This method provides evidence regarding how households in different quantiles respond to the audit program. The analysis focuses on overall survey participation, but we also report the results separately for web-based and mail-in program participants and the impact on consumption over time. The results suggest that there is a significant reduction

[9] Initially, we started the analysis by taking randomly selected customers (non-participants) as a comparison group. The confounding difference between the treatment and comparison groups is sufficiently convincing to not pursue this 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 participants.

in consumption overall with audit participation. Web-based survey participants show much greater reactions to their surveys than mail-in participants do (11% vs. 4.4%).[10] To test the durability of the intervention, short- (quarterly) and long-term (year) effects are also tested. Because DID and Quantile DID are being estimated, 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 are in the following sub-sections.

5.1. Graphical Results and Balance Diagnostics

The matching procedure was effective in creating a group of customers who were comparable with the treatment group concerning all of the observable confounders. 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. Then, we graphically display the distribution of propensity scores of the treatment and control groups (Figures 1A and 1B) 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 common support. In each graph, there is sufficient overlap between the treatment and control groups. This result suggests that with these similar groups, reasonable comparisons can be made. Then, each individual in the treatment group is matched with

[10] We also estimate whether difference in reactions to the survey between web and mail participants are statistically significant. In order to analyze that we employ "difference-in-difference-in-difference" (triple difference) method suggested by Hamermesh and Trejo (2000). The triple difference estimation shows statistical significance of the difference. These differences in response rate could also be attributed to the some difference in household characteristics between mail-in and online survey participants. See the Appendix for elaboration.

individuals in the comparison group based on the kernel-based propensity scores. Figures 1A and 1B 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 distribution. 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 results suggest that the matched groups' balance is off by only a small amount, where the values of the standardized bias for overall HEES participation is 0.5%, which is less than the unmatched maximum of 58.8%. Moreover, the differences between the groups became not statistically significant 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 14.1%. This result suggests that the group of online participants in particular (even before the matching) was more similar than the general survey and mail-in participants were. 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 35.5% to 5.8%. Studies suggest that the standardized bias should be less than 5%-10% (Rosenbaum and Rubin, 1985; Austin, 2009). In addition, the t-test is influenced by the sample size in the data (Austin, 2009). For mail-in participants, the size of future participants is much greater than the treatment group, and greater emphasis should not be placed on the t-test than on the standardized percent bias.

5.2. Estimation Results

This sub-section examines various measures over a 2-year period to investigate how customers who participated in the energy efficiency audits perform, on average (individual and distribution), compared with customers who waited one year to participate. 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 outcomes of the DID estimations, where the outcome is the natural log of electricity consumption. Columns 1, 2, and 3 show the DID estimations without matching, and columns 4, 5, and 6 show the propensity score estimations. The significance of the coefficients, the small differences among the coefficients (approximately 1 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 energy audit program participation, the customers who participated in the survey in January 2009 reduced their electricity consumption by 6.7%, or 75.57kWh, compared with households that did not participate in the survey until January 2010. The differential performance of online survey participation compared with mail-in survey participation is also important. Tables 3a and 3b show that on average, one entire year after HEES participation, the online HEES participants reduced their electricity consumption more than the mail-in participants, 11% vs. 4.4% (or 112 kWh vs. 52 kWh). Du et al. (2014) also show the same differential effect between online and mail-in HEES participants. In their study, the authors investigate the probability of future energy efficiency program participation as a function of current HEES participation. They conclude that the delivery mechanism of the survey matters for post-intervention behavior. Thus, the households that participated in the online survey increased the probability of future energy efficiency program participation by 3.4% to 4.3% compared to 2.6% for mail-in survey participation.[11] These two recent empirical results suggest that utilities and program designers could achieve greater behavioral responses of either 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- and long-term behavior. The frequency of the time investigated is quarterly. As discussed earlier, the HEES provides personalized feedback and energy conservation information. This survey does not provide repeated interaction, as in other home energy reports such as Opower energy reports. Therefore, we are also interested in how customers respond to HEES audit programs in the months or year after the surveys. Allcott and Rogers' (2014) Opower study suggests that there is an immediate response to the initial reports. 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. Table 4a shows that in contrast, households did not immediately respond to the non-binding personalized feedback. The average treatment

[11] 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.

effects increase gradually as time passes. There is no effect after the first three months. The effect after six months is 3.1%, and it is approximately 6% nine months later. There is a 6.8% reduction in electricity consumption one year later compared with households that have not yet joined the program. One year later, the treatment behavior does not attenuate, but instead, habitual behavior change is observed. However, there are diminishing returns as time passes. If we evaluate these conclusions together with the results from Du et al. (2014), the contrasting results from Allcott and Rogers (2013) 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 4.4% and 5.6%, respectively. Households in this group 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. As the study by Accenture (2012) demonstrates, utility consumers in the U.S. only think about their electricity consumption nine minutes per year, and their attention and interaction increase when they receive a high bill. These results (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, respectively. As shown in Table 4b, which shows the combined effect, there are immediate reactions to the surveys after the first three months.

In the following quarterly frequencies, there are 1.6%, 4.7%, and 4.4% reductions in electricity consumption. However, the reactions actually decrease instead of increase at a decreasing level. As shown in Table 4c, online survey participants reduced their consumption by 6.9%, 10%, and 11% over time. It would be much more interesting if longer-range consumption data were available.

Thus far, we have discussed the average treatment effect of program participation and have described the average effect of a survey on the individual utility customer. However, because the dependent variable has a continuous distribution, an inspection of averages may not properly explain the changes in the distributions (Angrist and 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 approach shows how households at different quantiles react to personalized feedback and the differences among the corresponding quantiles. To accomplish this task, we employ quantile DID by using kernel-based propensity score matching estimation. This approach provides a comfortable framework for examining how the quantiles of energy consumption change in response to survey participation. Thus, this technique allows us to detect heteroskedasticity (Deaton, 2000).

Tables 5a and 5b provide the quantile DID estimators and the effects of survey participation on both the central (0.25, 0.5, 0.75) and extreme quantiles (0.1 and 0.9). The estimates show significant effects of audit participation compared with households that have not yet participated in the survey. The results suggest that further away from the lowest quantiles, the magnitude of the slope decreases (see Figure 2). The slopes of each quantile regression differ. The 10^{th} and 90^{th} percentiles of the distribution are further apart from one another in the responses to the energy conservation surveys. Households in the lowest quantile save approximately 8.2% 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. These three columns in Table 5a suggest that the lower consumption quantiles saved much more than the upper quantiles among the survey participants compared with the comparison group. The changes in the slopes for the online audit participants are lower than for the mail-in participants. Thus, the results in Table 5a have more important implications for policy makers and program designers than simply considering the average effect. These implications suggest that consumers who are in the lowest quantiles are inclined to have more dramatic reactions to non-binding energy conservation than consumers in the median or highest quantiles. The critical result of these quantile regressions is showing that the estimated survey effects differ by the level of pre-survey household consumption.

Finally, Table 6 provides the estimates of kernel-based propensity score matching DID and quantile DID for the mail-in and online participants, which were computed by using the bootstrap method. Bootstrapped standard errors were estimated by using the same regressions to further check the robustness and replicability of the results. We checked only the quantile of 0.5 for the QDID regressions. However, as shown in Table 6, the bootstrap method provides similar, almost identical results for all of the estimators.

6. Concluding Remarks and Policy Implications

Residential electricity consumption represents approximately 35% of California's total electricity demand. California is ranked 50th in per capita residential electricity use.[12] In 2010, per capita electricity use in California was 2,337 kWh per capita, whereas the U.S average is twice as much at 4,674 kWh.[13] California's electricity use per capita has been almost flat from 1973 to the present, whereas the average U.S. electricity use has increased by 50 percent (Costa and Kahn, 2010)[14]. One of the factors that contributes to this trend is that the government has mandated energy efficiency standards for household appliances. However, spending on electricity is closer to the national average because of higher electricity prices in California (RECS, 2009). Additionally, Davis (2008) argues that because household production is time-intensive, even large changes in energy efficiency [in durable goods] will have little impact on demand. For example, EIA (2015) indicates that in U.S. households from 1980 to 2009, electricity use increased significantly, by 17%, compared with other fuel types. This trend has prompted policy makers to further strengthen the investment in non-price interventions and energy efficiency programs to reduce both consumption levels and carbon emissions. "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, Newell, and Palmer, 2004). In our research we examine one of these statewide home energy efficiency surveys (HEES) and determine how well the program has worked in terms of saving energy.

In particular, we have evaluated the effects of energy efficiency surveys, which give households feedback on their consumption behavior and provide energy conservation tips. In addition, investigated the differential performance from the mail-in and online versions of the audit program. Because survey participation is voluntary, the

[12] http://apps1.eere.energy.gov/states/residential.cfm/state=CA?print
[13] http://www.energyalmanac.ca.gov
[14] Working paper.

matching of the propensity scores reduced the differences between the treatment and the comparison groups. We also estimated the effects over time. We calculated two estimators using DID matching methods, namely, the Average Treatment Effects and Quantile Treatment Effects.

We provide evidence that the customers who participated in the survey reduced their electricity consumption by 6.7%, or 75.57 kWh, compared with customers who had not yet participated in the survey. The differential effects of the different versions of the survey show similar significant treatment effects, at 11% and 4.4% for online and mail-in participants, respectively. The implication of this finding is that how the program is delivered matters as much as having a program. Du et al. (2014) report similar findings. These two studies suggest that online energy efficiency programs have more impact on subsequent behavior. In addition, the results also suggest that the effects become significant and increase in magnitude gradually over time but at a decreasing level. The program 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. The 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 exhaustively examine habit formation, "habits increase the marginal utility of engaging in an activity in the future" (Charness and Gneezy, 2009). A mere educational and informational approach will not incentivize a household because of insufficient salience in the market. Instead, this approach can reverse the effects of policy goals and lead to inertia in consumption behavior and investment decisions, similar to the theory that was discussed in section 2. 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 and Hsiaw (2014) suggest that some households may actually view energy efficiency surveys as a commitment device. Therefore, it is necessary to have additional interactions with households. Additionally, because the persistence of treatment has a spillover effect for the year after the intervention and leads customers to other energy efficiency programs, an assessment of cost effectiveness should also include these externalities (Allcott and Rogers, 2014).

Because of the heterogeneity on pre-treatment energy consumption, we then examined the quantile DID estimator. The results suggest that as the quantiles of the distributions increase, the effect (concerning the natural logarithm of consumption) of the program decreases (see Figure 2). The implications of these findings for public policy are straightforward. Households at the lower quantiles of usage are more responsive to energy audit programs than are households at higher percentiles. Better customer targeting based on these distributions would create significant savings, which would also improve the efficiency of the programs and possibly have more lasting effects. However, concerning kWh reduction, smaller percentage change reductions by households in the higher quantiles can achieve more savings of kWh (see Table 5b). These results suggest that public utilities should properly define the savings and their objectives. Furthermore, because of the nature of these observational studies, policy makers should emphasize a distributional evaluation approach over the average treatment effects on an individual. Additionally, insights from behavioral economics can also be effective in addressing behavioral failures. Thus, better targeted information can elicit biased belief, present bias and other decision biases (Allcott, 2014; Allcott and Taubinsky, 2015).

In implementing the method that suggested by Sianesi (2004, 2008), we determined the adequate comparison group to correct for the selection bias in non-experimental energy efficiency program evaluations. Next, we employed a diagnostic test that is similar to the method suggested by Austin (2009) after matching the estimated kernel-based propensity scores. Then, combined with the two regression estimators, we described ways to address the systematic differences between the treated and comparison individuals in investigation of the effects of residential energy efficiency surveys. Our research is unique in its application such combined methods in evaluating residential energy efficiency programs.

References

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

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

Allcott, H. 2014. Paternalism and Energy Efficiency: An Overview. NBER, Working Paper 20363

Allcott, H., and Mullainathan, S. 2010. Behavioral Science and Energy Policy. *Science* 327 (5970): 1204–1205.

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

Allcott, H., and Taubinsky, D. 2015. Evaluating behaviorally-motivated policies: Experimental evidence from the light bulb market. American Economic Review, forthcoming

Angrist, D. J., and Pischke, J-S. 2009. Mostly Harmless Econometrics. Princeton University Press. Princeton, NJ

Athey, S., and Imbens , W. G. 2006. Identification and Inference in Nonlinear Difference-in-Difference Models. Econometrica. 74(2): 431–497

Austin, C, P. 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, O. 2006. Too Much Ado about Propensity Score Models? Comparing Methods of Propensity Score Matching. Value in Health 9(6): 377–385

Black, N. 1996. Why We Need Observational Studies to Evaluate the Effectiveness of Health Care. *BMJ* . May 11.

Black, D., and Smith, J. 2003. How Robust is the Evidence on the Effects based on Propensity Scores. *The Stata Journal* 2(4): 358–377.

Brennan, J. 2010. Decoupling in Electric Utilities. *Journal of regulatory Economics , 38*, 49-69.

Caliendo, M., and Kopeinig, S. 2005. Some Practical Guidance for the Implementation of Propensity Score Matching. IZA DP No. 1588

Charness, G., and Gneezy, U. 2009. Incentives to Exercise. *Econometrica* 77(3): 909–931

Costa, D. L., and Kahn, E. M. 2010. Why Has California's Residential Electricity Consumption Been So Flat Since the 1980s? A Microeconomic Approach. NBER Working Paper 15978

Deaton, A. 2000. The Analysis of Household Surveys: A Microeconomics Approach to Development Policy. 3rd Ed. Published for the World Bank. The John Hopkins University Press. Baltimore.

Dehejia, H. and Wahba, S. 2002. Propensity Score-Matching Methods for Nonexperimental Causal Studies. *The Review of Economics and Statistics , 84*, 151–161.

DellaVigna, S. 2009. Psychology and Economics: Evidence from the Field. *Journal of Economic Literature,* 47, 2, 315–72.

Du, Y., Hanna, D., Shelton, J., and Buege, A. 2014. What Behaviors Do Behavior Programs Change. *ACEEE Summer Study on Energy Efficiency in Buildings*

Frondel, M., and Vance, C. 2013. Heterogeneity in the Effect of Home Energy Audits: Theory and Evidence. *Environmental and Resource Economics*, European Association of Environmental and Resource Economists, vol. 55(3), pages 407–418, July.

Fu, A., Dow, W., and Liu, 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.

Hamermesh, S. D., and Trejo, J. S. 2008. The Demand for Hours of Labor: Direct Evidence from California. *The Review of Economics and Statistics* 82(1): 38-47

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

Heckman, J., Ichimura, H., Smith, J., and Todd, P. 1998. Characterizing Selection Bias Using Experimental Data. *Econometrica* 66 (5) :1017–98.

Heckman, J., LaLonde, R., and Smith, J. 1999. The Economics and Econometrics of Active Labor Market Programs. *Handbook of Labor Economics Vol.III*, ed. By O. Ashenfelter, and D. Card, pp. 1865-2097. Elsevier, Amsterdam.

Imbens, W.G., and Rubin, B.D. 2015. Causal Inference for Statistics, Social, and Biomedical Sciences: An Introduction. Cambridge University Press, New York, NY.

Ingle, A., Moezzi, M., Lutzenhiser, L., Hathaway, Z., Lutzenhiser, S., Van Clock, J., Peters, J., Smith, R., Heslam, D., and Diamond, R.C. 2012. "Behavioral Perspectives on Home Energy Audits: The Role of Auditors, Labels, Reports, and Audit Tools on

Homeowner Decision-Making". *Lawrence Berkeley National Laboratory*.

Ito, K. 2012. Do Consumers Respond to Marginal or Average Price? Evidence from Nonlinear Electricity Pricing. Energy Institute at Haas Working Paper 210.

Gillingham, K., Newell, G.R., and Palmer, K. 2004. Retrospective Examination of Demand-Side Energy Efficiency Policies. RFF DP 04-19 REV

Kennedy, L. 2003. A Guide to Econometrics. 5th Edition. The MIT Press. Cambridge, MA.

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

List, A. J., Millimet, L., Fredriksson, G. P., and McHone, 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, D. B. 1995. Natural and Quasi-Experiments in Economics. *Journal of Business and Economic Statistics , 13* (2), 151–161.

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

Rosenbaum, R., and Rubin, B. 1983. The central role of the propensity score in observational studies for causal effects. *Biometrica , 70*, 41–55.

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

Rubin, B.D., and Thomas, N. 1992. Affinely Invariant Matching Methods with Ellipsoidal Distributions. *The Annals of Statistics.* 20(2): 1079–1093

Sexton, S. 2015. Automatic Bill Payment and Salience Effects: Evidence from Electricity Consumption. *The Review of Economics and Statistics* 97(2):229–241.

Shin, J.-S. 1985. Perception of Price When Price Information Is Costly: Evidence from Residential Electricity Demand. *The Review of Economics and Statistics* 67 (4): 591–98.

Sianesi, B. 2004. An evaluation of the Swedish system of active labour market programs in the 1990s. *The Review of Economics and Statistics* 1, 133–155.

Sianesi, B. 2008. Differential effects of active labour market programs for the unemployed. *Labour Economics* 15(2008) 370-399

Smith, H. 1997. Matching with Multiple Controls to Estimate Treatment Effects in Observational Studies. Sociological Methodology 27:325-353

Titus, A. M. 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. 2009. Residential Energy Consumption Survey. Retrieved March 19 from http://www.eia.gov/consumption/residential/

Wirl, F. and Orasch, W. 1998. Analysis of United States' Utility Conservation Programs. *Review of Industrial Organization* 13: 467 – 486

Zwane, A. P., Zinman, J., Van Dusen, E., Parienta, W., Null C., Miguel, E., Kremer M., Karlan D. S., Hornbeck R., Gine X., Duflo E., Devoto F., Crepon B., and Banerjee, A. 2011. Being Surveyed Can Change Later Behavior and Related Parameter Estimates. *PNAS* 108(5): 1821-1826.

Figures and Tables[15]

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 (0.7)	6.89 (0.71)	6.78 (0.65)

Note: Standard deviations are in parentheses. Percentages are rounded. 97.48 % of the households in the data have 24 months of observation, 2.52% varies between 15 – 23 months.

Table 2: Balance diagnostics across all the estimated propensity scores

VARIABLE	Unmatched	Mean		% bias	t - test	
	Matched	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.

[15] All models control for a household billing, demographics, dwelling characteristics, survey data and as well as weather variables.

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.

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-squared	0.393	0.322	0.365	0.358	0.360	0.371

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

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.

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-squared	0.2730	0.250	0.350	0.269	0.204	0.362

Note: Standard errors in the 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. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

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. ***Aggregate / Combined survey participation.***

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)
Constant	6.602***	6.611***	6.609***	6.604***
	(0.0165)	(0.0152)	(0.0144)	(0.0136)
Observations	52,320	62,825	73,323	83,834
R-squared	0.387	0.388	0.387	0.384

Note: Standard errors 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 – entire period. *** p < 0.01, ** p < 0.05, * p <0.

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 participation*.**

VARIABLES	(1) Ln(KWh)	(2) Ln(KWh)	(3) Ln(KWh)	(4) Ln(KWh)
Diff-in-Diff - M	-0.000240	-0.0158**	-0.0458***	-0.0438***
	(0.0107)	(0.00746)	(0.00648)	(0.00591)
Constant	6.343***	6.342***	6.416***	6.436***
	(0.0179)	(0.0167)	(0.0186)	(0.0154)
Observations	43,451	52,343	60,780	69,514
R-squared	0.362	0.362	0.365	0.360

Note: Standard errors 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

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 version of the survey participation.*** Dependent Variable: Log Consumption.

VARIABLES	(1) Ln(KWh)	(2) Ln(KWh)	(3) Ln(KWh)	(4) Ln(KWh)
Diff-in-Diff - O	-0.0266	-0.0670***	-0.0964***	-0.110***
	(0.0268)	(0.0209)	(0.0190)	(0.0175)
Constant	6.372***	6.347***	6.331***	6.327***
	(0.0477)	(0.0482)	(0.0460)	(0.0435)
Observations	9,372	11,278	13,205	15,131
R-squared	0.369	0.375	0.372	0.371

Note: Standard errors 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). *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

Table 5a: Kernel propensity score matching Quantile DID 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 in parentheses. Estimations are adjusted for seasonality using seasonal dummies.
*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

Table 5b: Kernel propensity score matching Quantile DID estimation of the *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 in parentheses. Estimations are adjusted for seasonality using seasonal dummies.
*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

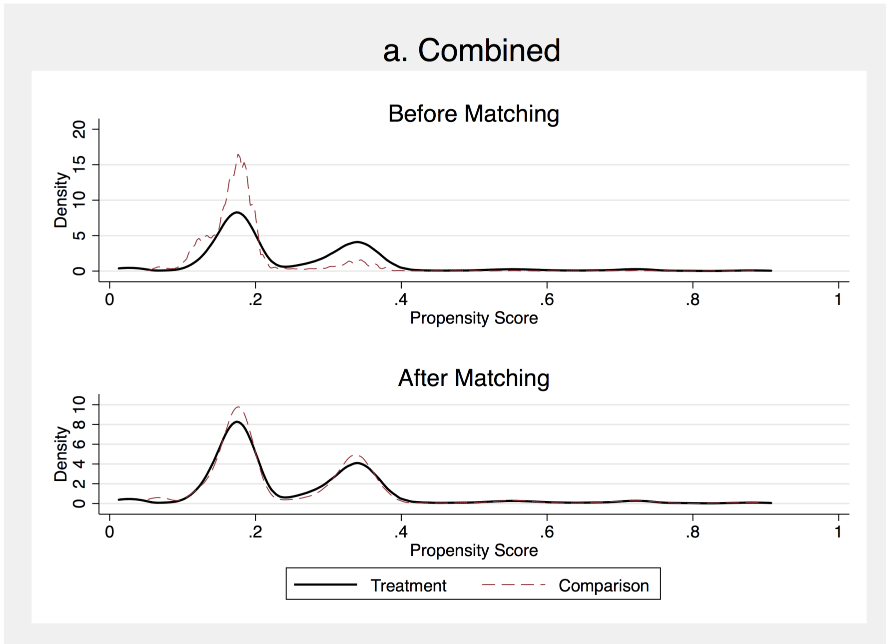
Table 6: (*Mail-in and Online sample*) bootstrapped coefficients and standard errors (100 repetitions), Kernel Propensity score matching DID and Quantile (q=0.5) DID.

	(1)	(2)	(3)
VARIABLES	Ln(kWh)	kWh	kWh (q = 0.5)
B_*Diff-in-diff* - M	-0.0438***	-51.91***	-39.41***
	(0.00984)	(18.69)	(10.23)
B_*Diff-in-diff* - O	-0.110***	-112.4***	-61.91***
	(0.0167)	(19.34)	(12.61)
Observations - M	69,514	69,526	70,454
Observations - O	15,131	15,241	15,272

Note: Standard errors in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

Figure 1. **Estimated propensity of scores, by groups.** Pre- and post-matching density estimates of propensity scores among the treatment and comparison groups (Epanechnikov kernel, the bandwidth is 0.06 - default).

A)



B)

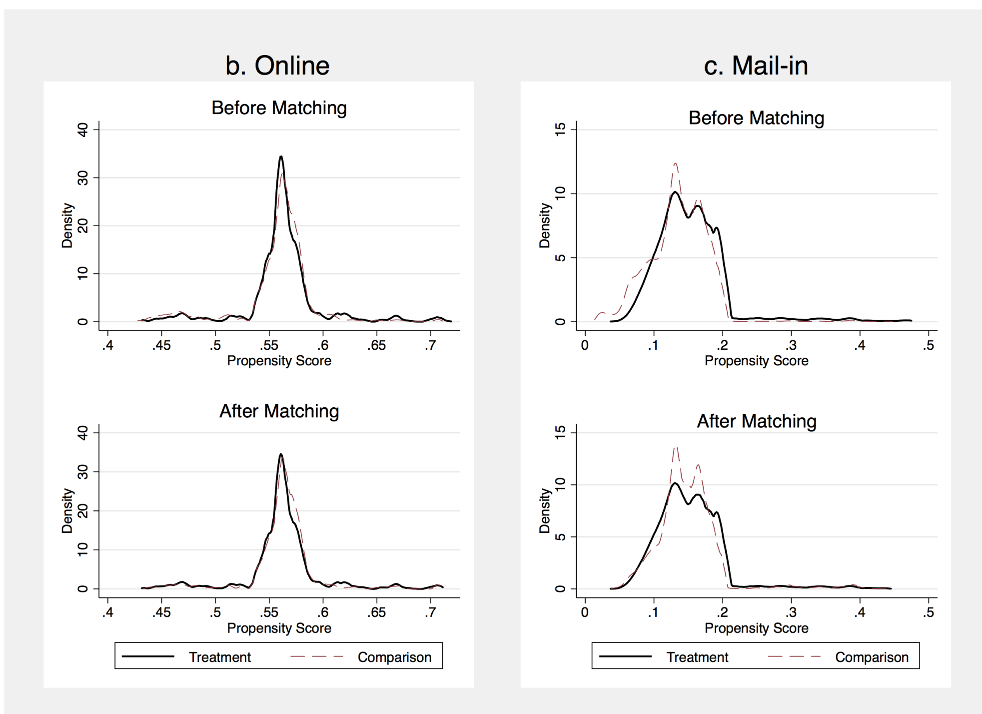
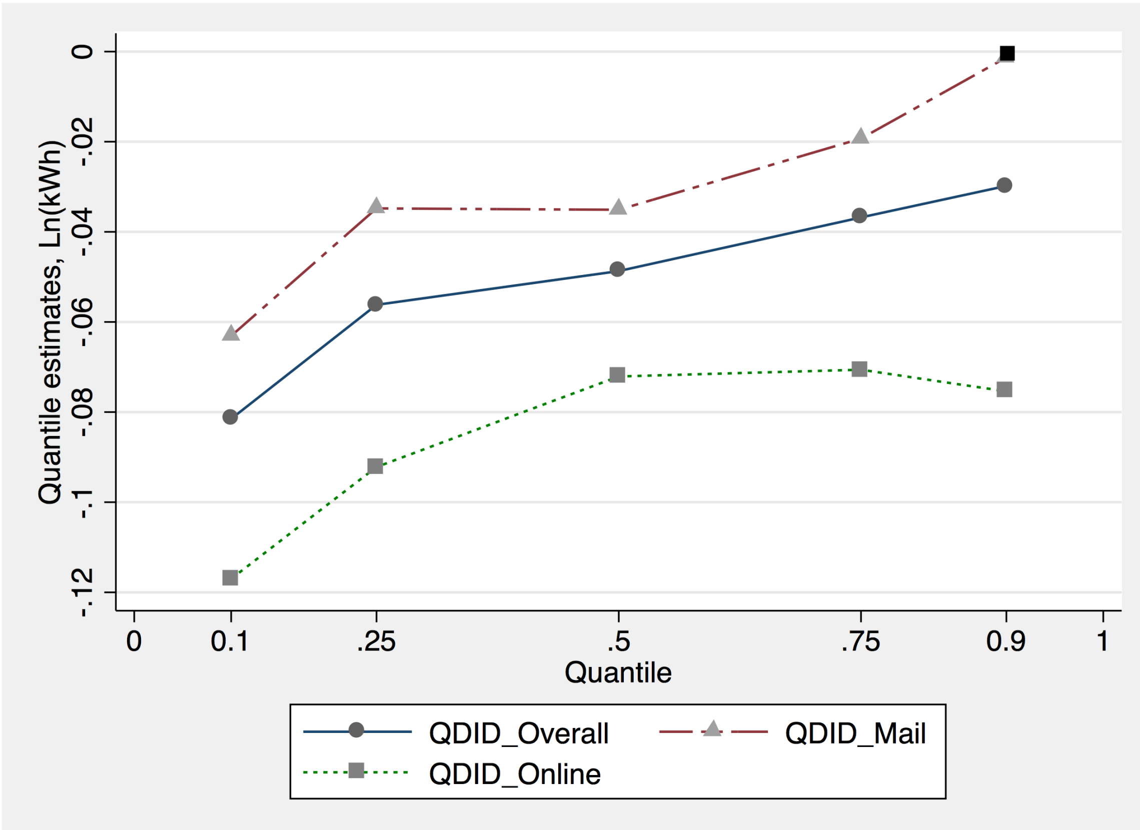


Figure 2: Matching Quantile DID estimates of efficiency program participation effects. Dependent Variable: Log Consumption



Note: At the 90th percentile: The coefficient of QDID_Mail is not significant.

APPENDIX

Variable Name	Description / range of values	Mean	St.D.
kWh Usage	Monthly Electricity consumption in kWh data from 2008 January – 2009 December for 4173 randomly selected customers from all income groups (99568).	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-\$15000, maximum is \$250,000. In this study, incomes greater than \$100,000 merged in single category	7.23	
Household Size (Range 1-8)	Number of people in the household. Values range from 1-8	3.59	1.88
Rooms	Number of room in the house. Values range from 1-8	7.9	2.9
Square Foot	Size of the house, in square foot.	2480	1076
CDD	Cooling Degree Days (at 72F)	77	111.3

	Survey Types - Household Characteristics	
Variables	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%

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.