Energy Economics 52 (2015) S30–S40

Contents lists available at ScienceDirect

Energy Economics

journal homepage: www.elsevier.com/locate/eneeco

Information v. energy efficiency incentives: Evidence from residential electricity consumption in Maryland☆

Anna Alberini [a,b], Charles Towe [c]

[a] Department of Agricultural and Resource Economics, University of Maryland, USA
[b] Centre for Energy Policy and Economics, ETH Zürich, Switzerland
[c] Department of Agricultural and Resource Economics, University of Connecticut, USA

ARTICLE INFO

Available online 28 August 2015

JEL classification:
Q41
D12
H3

Keywords:
Energy efficiency
Household behavior
Energy efficiency incentives
Electricity usage
Home energy audit

ABSTRACT

We focus on two utility programs intended to reduce energy usage and the associated CO2 emissions—a home energy audit and rebates on the purchase of high-efficiency air-source heat pumps. We use a unique panel dataset from participating and non-participating households to estimate the average treatment effect of participating in either program on electricity usage. We fit models with household-by-season, season-by-year, and household-by-year fixed effects to account for all possible confounders that might influence energy usage. Since the programs are voluntary, we seek to restore near-exogeneity of the program "treatment" by matching participating households with control households. We deploy coarsened exact matching (CEM; Iacus et al., 2011) as our main matching method. We ask whether it is sufficient to match households based on past electricity usage, or if we gain by adding structural characteristics of the home, including heating system type. We find that the two programs reduce electricity usage by 5% on average. The effects are strong in both winter and summer for the energy audit groups but appear to be stronger in the winter for the heat pump rebate group. Adding house characteristics to the matching variables does seem to affect results, suggesting that using past usage alone, or house characteristics alone, may not be sufficient to identify the effects of program participation.

1. Introduction

The purpose of this paper is to estimate the savings in residential electricity usage that can be attributed to energy efficiency programs. We focus on two such programs. The first is a home energy audit offered to customers free of charge, where information is provided to the consumer about ways to save energy and money, and the consumer is free to choose which advice to implement, and when. The second is a rebate on the purchase of a high-efficiency heat pump, a device used for heating the home in the winter and cooling it in the summer that is very common in our study area due to its climatic conditions (hot summers and winters marked with only brief exposure to extreme cold).

The two programs rely on completely different approaches to encouraging energy efficiency investments in the home: The former provides information at low or no cost to the consumer, while the latter lowers the capital cost of the investment. We interpret participation in either of these two programs as a "treatment" in the context of an experiment with residential electricity consumption as the outcome. We assess the effect of the treatment using a unique panel of data on electricity usage before and after the time of the program for both the participating households and suitable control households.

Environmental issues and climate change concerns have led to a resurgence of residential energy efficiency programs by policymakers and utilities wishing to reduce energy usage and the CO2 emissions associated with electricity generation. Well publicized and influential reports (Intergovernmental Panel on Climate Change, 2007; McKinsey and Company, 2009) have identified energy efficiency improvements in buildings as capable of delivering CO2 emissions reductions at low or even negative cost, and in the US in fiscal year 2013 federal expenditures on preferential tax policies targeting energy efficiency improvements in existing and new homes came to a total of almost $4 billion (Dinan, 2013).

Residential efficiency programs were popular among the utilities in the late 1980s and early 1990s, when they were part of the utilities' demand-side management programs, which attempted to reduce electricity usage to avoid or postpone expensive capital expenditures and reduce peak load. However, the cost-effectiveness of these measures was and still is difficult to study, due to adverse selection and the likelihood that these programs attract people who are systematically (and unobservably) more motivated or productive at reducing electricity usage (Allaire and Brown, 2012; Hartman, 1988; Joskow and Marron, 1992; Waldman and Ozog, 1996). As a result, considerable debate

☆ We thank the attendees of the 6th Atlantic Workshop on Energy and Environmental Economics in Atoxa, Spain, and of a seminar at KU Leuven, Brussels, for their helpful comments.

E-mail addresses: aalberin@umd.edu (A. Alberini), charles.towe@uconn.edu (C. Towe).

http://dx.doi.org/10.1016/j.eneco.2015.08.013
0140-9883/

remains about the cost-effectiveness of these programs (Auffhammer et al., 2008; Loughran and Kulick, 2004).

Ideally, one would want to evaluate residential energy efficiency programs by conducting randomized controlled trials, where households are exogenously assigned to treatments of different type or intensity (Davis, 2008). Alternatively, it might be possible to devise circumstances that are plausibly interpreted as natural experiments (Gans et al., 2013). Our study lends itself to neither of these criteria nor do we have plausible instruments for participation. Fortunately, we do have the data necessary to implement a retrospective case–control study and address these problems. We capture all confounders through a "triple difference" approach that lets log household electricity usage depend on the weather, household-by-season fixed effects, season-by-year fixed effects, and household-by-year fixed effects. This setup takes advantage of the panel nature of our dataset.

We also match treated households with similar control households, based on (i) electricity usage in the benchmark year (2008), (ii) structural characteristics of the home, and (ii) both (i) and (ii). By design and when properly implemented the matching procedure restores balance across treatment and control households conditional on our rich set of observables and in theory makes treatment as good as randomly assigned—a so-called "quasi experiment".

One question we address in this paper is whether it is sufficient to match treated and control observations on past usage (usage during 2008, our benchmark year) or we gain by creating matching strata based on past usage *and* structural characteristics of the dwelling. Clearly, the former approach is coarser and results in more numerous matches (and hence a larger sample size), while the latter is more precise, but requires information beyond the mere usage history of the households, and by design may discard many more units.

If past electricity usage is sufficient to describe completely a household's energy usage patterns, then adding house characteristics should not make much difference in terms of the results of the matching exercise. If, on the other hand, structural characteristics of the dwelling help explain usage and/or participation in the utility program and contribute to covariate imbalance, then the matching exercise will give different results when we add the structural characteristics. Indeed, house characteristics are important determinants of energy demand (e.g., Alberini et al., 2011), so the matching exercise that relies on both house characteristics and benchmark usage might be thought of as the closest we can get to controlling for both observables (the house characteristics, which could be use to predict electricity demand) and unobservables, such as taste for a warm or cool home (which would be absorbed into the benchmark usage, and might be interpreted as the portion of usage that is not predicted by house characteristics).

We do not have a priori expectations on whether the two approaches produce very different results, but we note that recent literature about the effects of novel tariffs or utility program has often relied on just past usage, with no information about the structural characteristics of the dwelling (Auffhammer, 2014).

We use coarsened exact matching (Iacus et al., 2011) to match households, and apply the resulting weights in regressions that use the full panel of observations. We compare the results from this approach with standard matching (Abadie and Imbens, 2006, 2011) and propensity score matching (Dehejia and Wahba, 1999, 2002) for cross-sections drawn from our full sample.

Briefly, we find that the energy audit and rebates on the purchase of high-efficiency air-source heat pumps resulted in 5% reductions in the use of the electricity. The savings appear to be equally strong in the winter and summer in the case of the energy audits, and the results are sharper when matching is done on both past usage and house characteristics, despite the considerable trimming the sample is subjected to. With the heat pump rebates, the savings accrue primarily in the winter. In sum, our results suggest that matching on only past usage may not be enough, and that usage data should be augmented with house and/or household characteristics when possible.

Our paper is different from recent work in the area of "information" about energy usage, which has focused on examining whether more frequent feedback on usage than the conventional billing frequency, simplified or reformulated bills, or real-time feedback on usage through in-home displays (alone or combined with dynamic pricing) change household energy consumption (Faruqui et al., 2010; Gans et al., 2013; Jessoe and Rapson, 2014). We contribute to the strand of literature that has sought to assess energy-efficiency incentive programs by examining the uptake of such incentives (Hassett and Metcalf, 1995), free riding in their presence (Boomhower and Davis, 2014) and apparent rebound effects potentially induced by the availability of these incentives (Alberini et al., 2014).

The remainder of this paper is organized as follows. Section 2 presents the background for our study. Section 3 describes the data. Section 4 lays out the econometric model and methods. Section 5 presents the results and Section 6 offers concluding remarks.

2. Background

In 2008, the state of Maryland established the EmPower Maryland Program, with the goal of reducing energy consumption by 15% by 2015. Participating electric and gas utilities set up a number of initiatives to help meet this goal, including—starting in January 2010—rebates of $200 and $400 on the purchase of air-source heat pumps in tier I and tier II, respectively. This rebate structure remained in place for all of 2010 and 2011, and was revised in January 2012, when rebates were extended to tier III heat pumps and ductless mini-split heat pumps that met specific energy efficiency requirements. The electric utility that serves the study area is a participant in the EmPower Maryland program.

In January 2011 the participating utilities started home energy audit programs. In this paper we examine the effects on energy usage of the simplest and least time-consuming of these audits—the Quick Home Energy Check-up (QHEC). In the QHEC, a professional performs a one-hour walk through the home to assess insulation levels, air leakage, heating and cooling systems, windows and doors, lighting and appliances, and water heating equipment. A report is prepared and handed to the homeowner that summarizes findings and recommends improvements and opportunities to save energy use and costs. Equipment and supplies, such as compact fluorescent light bulbs, faucet aerators, efficient-flow showerheads, water pipe insulation or water heater tank wraps, are offered. The QHEC is free to the residential customer and costs about $200 to the utility (which employs a contractor to do this service).

We do not know exactly what a household does after the free energy audit. It is possible that, in addition to accepting and installing the products offered at the time of the QHEC, the audited households replace major equipment or install insulation, but we do not know this.

We wish to assess the effect on electricity usage of participation in the heat pump rebate program or the QHEC in first quarter of 2011. As mentioned, the goal of the rebate and the QHEC program is to help the utilities meet the requirements of the EmPower Maryland program, which in turn aims at a 15% reduction in energy use and at the associated CO_2 emissions. The utilities fund these programs by applying a surcharge to all residential customers, whether or not they partake in the programs. The surcharge is currently some 0.4 cents per kWh, which accounts for a negligible share of the total price per kWh paid by consumers (9–11 cents per kWh).[1]

Additional incentives have been available from the federal government in the form of tax credits on the purchase of high-efficiency heat pumps since 2006, with major revisions to tax credits and caps in 2009 as part of the American Reinvestment and Recovery Act. The

[1] For a residential customer with monthly usage approximately equal to the average in the US (1000 kWh/month), this implies an additional cost of $4 per month.

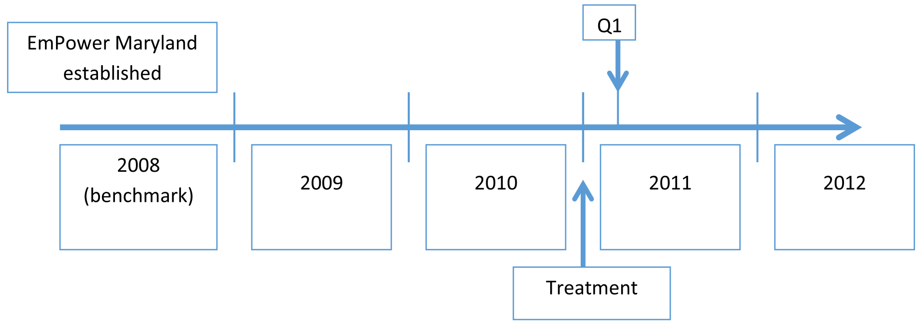

Fig. 1. Time line.

Table 1
Electricity consumption by year and household group.

	N	Mean electricity usage in 2008 (benchmark year) (kWh)	Mean electricity usage in 2009 (kWh)	Mean electricity usage in 2010 (kWh)	Mean electricity usage in 2011 (kWh)	Mean electricity usage Q2 2011–Q1 2012 (kWh)
Audit (QHEC)	378	17,438	18,046	19.049	17,640	16,407
Control group (control group (a))	10,676	17,385	17,652	17,853	17,970	16,846
Heat pump rebate	430	20,136	21,154	21,950	19,775	18,222
Control group with heat pump (control group (b))	6645	18,586	18,893	18,963	18,902	17,479

combination of utility rebates and federal tax credits helps defray the cost of a new heat pump, which ranges between $2000 and $20,000, depending on whether ductwork is required (ductless mini-split require no ductwork) and on the capacity of the system. No federal tax credits are available, however, for a home energy audit.

3. The data

We have assembled a unique dataset from state and private sources which contains monthly electricity usage and bills for a sample of about 17,000 households in Maryland. This sample is comprised of households who received a Quick Home Energy Check-up (QHEC) or a rebate for an energy-efficient air-source heat pump in the first quarter of 2011 (Q1 2011), plus households living in homes that are representative for age and construction type of the stock of single-family homes and townhomes in the area served by the utility, but did not participate in any utility programs during our study period.

Although the local utility provided us with monthly billing and usage information from as early as December 2006, in this paper attention is restricted to 2008–2012. Specifically, we use observations from 2008 for benchmarking purposes, and 2009 and the later years for analysis purposes (see the time line in Fig. 1). We use 2008 for benchmarking as this period was one of little or no programmatic activity targeted at energy efficiency. Since we are interested in assessing the effect of audits and incentives towards a major heating and cooling device, we exclude from the sample households that received multiple incentives during our study period or during any period after the first quarter of 2011. We include in our sample only households with accounts that were active in 2008 and remained active until at least Q2 2011 at the same home. Our cleaned sample is thus comprised of 378 QHEC households, 430 households who received a rebate on the purchase of air-source heat pumps with SEER of 14 or better, and 10,676 "control" households.[2] A total of 6645 out of these 10,676 households live in homes served by air-source heat pumps.

[2] The original list provided by the utility contained a total of 1300 households who received a QHEC or a heat pump rebate in Q1 2011, so our data cleaning procedures drop about one-third of the original households that participated in these utility programs in Q1 2011.

Information about electricity usage for this cleaned sample is displayed in Table 1. Annual average consumption in 2008, our "benchmark" year, ranges from 17,000 to over 20,000 kWh. This figure is above the US average (which is about 11,000 kWh[3]), in part because of the reliance on air conditioning in the summer in our study area and because over half of the homes in our sample are served by air-source heat pumps, which are heavy users of electricity. T tests (reported in Table 2) fail to reject the null that the audit and the control households have different mean consumption levels in 2008, and find that rebate recipients are significantly different from the full control group (control group (a)) and those in the control group that use heat pumps as their main heating and cooling system (control group (b)).

The distributions of electricity usage in 2008 for the different groups of households are depicted in Figs. 2 and 3. The figures suggest that, after some trimming at the upper end of the distribution, there is a wide common support for 2008 usage for treatment and control households.

Table 1 also reports information about electricity usage in 2009, 2010, 2011, and Q2 2011–Q1 2012, i.e., the twelve months after the utility programs. Usage appears to be especially high in 2009 and 2010 among the recipients of the heat pump rebates, and appears to decline substantially thereafter. Control households that use heat pump experience a comparatively much more modest decline in usage.

For each of the homes in our sample, we have extensive information about the structural characteristics of the dwelling and the type of heating and cooling system. This information comes from MDPropertyView, a database compiled by the State of Maryland that documents all properties in the state. While information about the house is extensive, we have no information about the socioeconomic characteristics of the households that live in the homes in our sample.

Descriptive statistics of selected housing characteristics from MDPropertyView are displayed in Table 3. Briefly, the first column of Table 3 shows that the average home is about 1900 square feet and that 62% of the homes use heat pumps as their main heating and cooling systems. The bulk of the homes in our sample—some 60%—were built in the 1980s and 1990s, and a majority (over 54%) are classified as of "average" construction quality. We note that higher construction quality includes "tighter" homes with regard to energy efficiency.

[3] See http://www.eia.gov/tools/faqs/faq.cfm?id=97&t=3 (last accessed 26 December 2014).

Table 2
2008 electricity usage comparison across groups.

	T statistic of the null that the group means are the same
Audit (QHEC) v. control group (a)	−0.1065
Heat pump rebate v. control group (a)	−6.9494
Heat pump rebate v. control group with heat pumps (control group (b))	−3.8586

In the remaining columns of Table 3 we compare the structural characteristics of the homes across groups—the two treated groups, the full control group (control group (a)), and the subset of the control group that use heat pumps (control group (b)). This comparison suggests that the QHEC group and the full control group are reasonably similar to each other, as are the heat pump rebate group and the controls with heat pumps. Some differences exist, however, in terms of the share of relatively new homes, construction quality, presence of basement and construction techniques and materials.

Finally, we use the daily average temperature in our study area from the National Climatic Data Center's Global Summary of the Day to compute daily heating and cooling degree days (HDDs and CDDs, respectively). Since the weather is a major determinant of the demand for electricity, we aggregate daily HDDs and CDDs to the seasonal totals for each household and enter them in the right-hand side of our regressions.

4. Econometric approach

4.1. The model

We are interested in assessing the effect of two alternate treatments, the energy audit and the rebate on the purchase of an efficient heat pump, on electricity consumption. We focus on households that received the energy audit, or received and redeemed the rebate for a new heat pump, in Q1 2011. We have their electricity consumption before and after Q1 2011, but do not know exactly when the audit took place or the heat pump was installed within the first quarter of 2011. For this reason, we aggregate the monthly electricity usage records to seasonal totals, and in our estimations (described below) we exclude the observations from Q1 2011. Electricity consumption is likewise aggregated to seasonal totals over the same study period for the control subjects. We define the seasons as

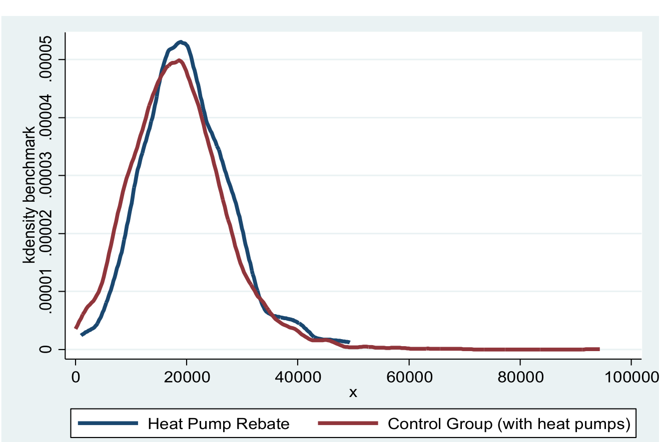

Fig. 2. Density of 2008 electricity usage. Households who received a heat pump rebate v. control households with heat pumps.

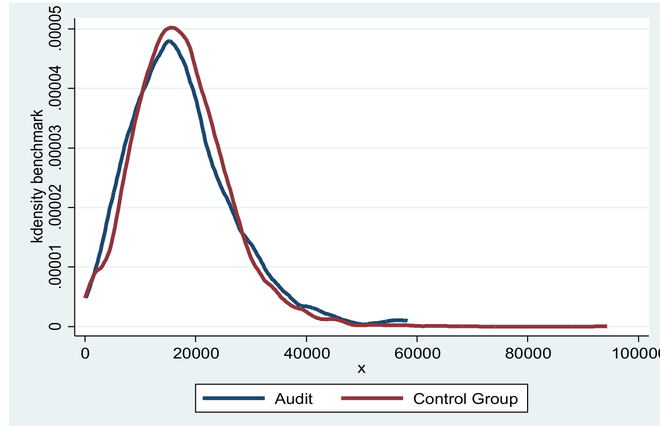

Fig. 3. Density of 2008 electricity usage. Households who received a Quick Home Energy Check-up (QHEC) v. control households.

winter (season 1), which is comprised of December, January, February, and March, spring (April and May), summer (June, July, August and September), and fall (October and November). In our study region, electricity consumption is especially high in the winter and the summer (even if we account for the different lengths of these seasons compared to spring and fall). This pattern is clear in Figs. 4 and 5, which display average log seasonal electricity use by customer group.

To control for unobserved confounders, we estimate the following "difference-in-difference-in-difference" equation:

$$\ln E_{ist} = \alpha_{is} + \tau_{st} + \theta_{it} + \mathbf{W}_{ist}\beta + D_{ist} \cdot \gamma + \varepsilon_{ist}, \tag{1}$$

where E is household's i electricity usage in season s in year t, α_{is} denotes a household-by-season fixed effect, τ_{st} a season-by-year fixed effect, and θ_{it} a household-by-year fixed effect. $\mathbf{W}_{ist}$ is a vector of weather controls, and D_{ist} is the treatment dummy. We are especially interested in estimating γ, the average treatment effect on the treated (ATT).

The household-by-season fixed effects capture preferences for a warm house in the winter and a cool house in the summer, insulation and ventilation characteristics of the home, the presence of tree shade, etc. The season-by-year fixed effects capture the shocks represented by unusually cold or warm winters or summers, and the household-by-year fixed effects any changes in the composition of the household or structural characteristics of the home from one year to the next that may influence electricity usage.[4] The effect of the treatment is identified by variation within the household–season–year cell.

In practice, Eq. (1) implies a large number of household fixed effects—a total of 16 effects per household times the over 10,000 households. Estimation is simplified by first taking the fourth-lag difference, namely the difference between each observation and its counterpart from the same season one year earlier. This removes the household-by-season fixed effects and yields

$$\ln E_{ist} - \ln E_{is,t-1} = \tau'_{st} + \theta'_{it} + (W_{ist} - W_{is,t-1})\beta + (D_{ist} - D_{is,t-1}) \cdot \gamma + e_{ist}, \tag{2}$$

[4] The household-by-season and the season-by-year fixed effects also account for the different lengths of winter and summer compared to the other seasons. The inclusion of household-by-season and household-by-year fixed effects allows us to relax the "common trends" assumption implicit in a difference-in-difference (DD) design and associated regression model, and accommodates for the treated households' heavier electricity usage in the two winters before Q1 2011.

Table 3
Descriptive statistics of the sample and comparison of structural characteristics of the homes by group: means of selected variables.

	All households (n = 11,552)	Audit	Control group (a)	Group means *t* test (audit v. control group)	HP rebate	Control group w/ heat pump (b)	Group means *t* test (HP rebate v. control group w/ heat pumps)
Square footage of the home	1928.19	1917.95	1931.22	−0.33	2072.4	2029.01	1.07
Heat pump (dummy)	0.6225	0.6085	0.6224	−0.52	1.0000	1.0000	–
Basement (dummy)	0.4209	0.3052	0.4285	−5.02**	0.4332	0.4582	−0.90
Frame (dummy)	0.0828	0.0736	0.1238	−3.59**	0.1217	0.0908	1.70
Brick (dummy)	0.1229	0.1062	0.0832	1.41	0.0534	0.0420	0.91
Siding (dummy)	0.7337	0.7411	0.731	0.43	0.813	0.8413	−1.30
Built Pre 1960 (dummy)	0.0594	0.0763	0.0612	1.07	0.0059	0.0104	−1.03
Built 1960–69 (dummy)	0.0649	0.0518	0.0665	1.25	0.0059	0.0051	0.20
Built 1970–79 (dummy)	0.1572	0.2071	0.1565	2.36*	0.0356	0.0508	−1.45
Built 1980–89 (dummy)	0.2170	0.2507	0.2124	1.67	0.3145	0.2735	1.58
Built 1990–2000 (dummy)	0.3921	0.3787	0.3879	−0.36	0.5964	0.5117	3.08**
Built 2000 + (dummy)	0.1094	0.0354	0.1155	−7.89**	0.0415	0.1485	−9.10**
1 floor (dummy)	0.3198	0.3578	0.3540	0.14	0.2614	0.2461	0.60
2 floors (dummy)	0.5583	0.6147	0.6236	−0.33	0.732	0.7205	0.44
2.5 floors (dummy)	0.0065	0.0122	0.0074	0.79	0.0033	0.0103	−1.98*
3 floors (dummy)	0.0132	0.0153	0.0150	0.04	0.0033	0.0232	−5.23**
Average construction quality (dummy)	0.5414	0.4796	0.5404	−2.29*	0.6291	0.6113	0.66
Fair construction quality (dummy)	0.3544	0.4142	0.3559	2.23*	0.2463	0.2721	1.07
Good construction quality (dummy)	0.0776	0.0899	0.0770	0.85	0.1128	0.1011	0.66
Very good construction quality (dummy)	0.0088	0.0054	0.0092	−0.94	0.0089	0.0133	0.84

* Significant at the 5% level.
** Significant at the 1% level.

where τ_{st}' and θ_{it}' denote new season-by-year and household-by-year fixed effects.

In certain runs, as when the sample is restricted to the summer just before and that just after participation in the utility program, the fourth lag difference results in a single observation per household. It is therefore not possible to fit a model with household-by-year fixed effects, and we estimate a simplified version of the "triple difference" model, namely:

$$\ln E_{ist} - \ln E_{is,t-1} = \tau'_{st} + (W_{ist} - W_{is,t-1})\beta + (D_{ist} - D_{is,t-1}) \cdot \gamma + e_{ist}. \quad (3)$$

While the interactions between the household, season and year units should help capture unobserved heterogeneity, Eq. (1) is linear in the logs of the continuous variables and the treatment dummy, which means that the model relies on extrapolation if certain cells are sparsely populated or are imbalanced with respect to the treatment and control households. To circumvent this problem, we deploy matching techniques in order to restore balance and near or plausible exogeneity of the treatment.

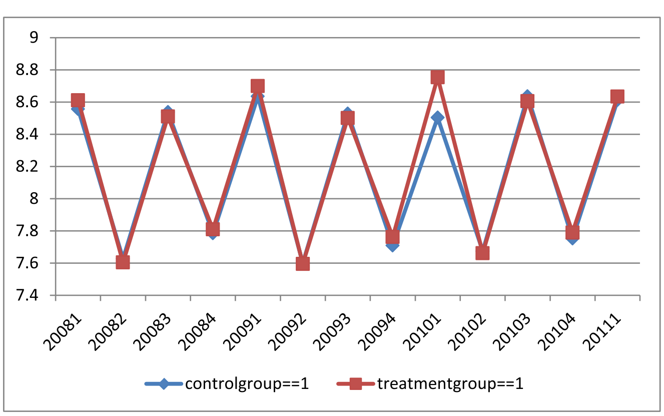

Fig. 4. Average log electricity consumption by season. Household who received a Quick Home Energy Check-up (energy audit) v. control households.

4.2. Matching

For each treated household, we look for a match, namely a control household with roughly the same levels of electricity usage in 2008 and/or similar dwelling characteristics. The simplest way to estimate the ATT is to compute the difference between log usage for each treated household and its control-group match, and then average these differences over all possible pairs of matched households.

We remind the reader that average treatment effect is defined as

$$\gamma^{ATT} = E(Y_1 - Y_0 | D = 1) = E(Y_1 | D = 1) - E(Y_0 | D = 1) \quad (4)$$

where Y_1 denotes the outcome for a household in the treated state, Y_0 denotes the outcome in the untreated state, and D indicates treatment status—in our case either participation in the audit or rebate program (Angrist and Pischke, 2009). Of course, we cannot observe $E(Y_0|D = 1)$ (the untreated outcome for treated households) which leads to utilizing data from the $D = 0$ group to estimate $E(Y_0|D = 0)$. All matching estimators of the ATT are weighting estimators of the form

$$\Delta = \frac{1}{n_1} \sum_{i \in \{D_i = 1\}} \left[Y_{1i} - \sum_{j \in \{D_j = 0\}} w(i, j) Y_{0j} \right] \quad (5)$$

where $w(i, j)$ sums to 1 for all i. These estimators assume non-confoundedness, which means we can simply construct matches without concern for selection into the treatment. Application of matching estimators in observational data requires addressing this selection issue conditional on a rich set of observables, **X**, which in our case include historical usage and dwelling characteristics.[5]

If the matching covariates **X** are solely binary indicators or categorical variables, then it is straightforward to construct strata defined by all possible combinations of **X** values and place the treated households and the controls in the appropriate stratum. The control households in the same stratum as any given treated households serve as matches for

[5] Matching relies on three key assumptions (Abadie and Imbens, 2006, 2011). The first is non-confoundedness, which means that treatment is exogenously assigned, conditional on the covariates **X**. The second is overlap, which states that for some $\eta > 0$, the probability of receiving treatment, conditional on the covariates **X**, is comprised between η and $(1 - \eta)$ (i.e., given **X**, we can find both treated and untreated units). The third is that the data points are independent draws from the distribution of the outcome variable, the treatment, and **X**.

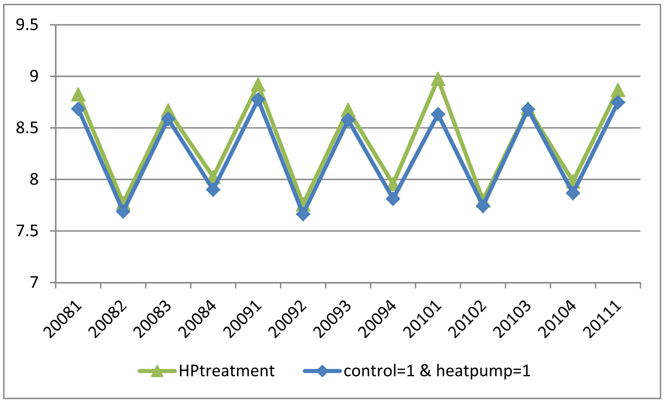

Fig. 5. Average log electricity consumption by season. Household who received heat pump rebates v. control households with heat pumps.

the latter. Under mild assumptions, the ATT in (5) is consistent and asymptotically normally distributed.

The inability to match for each continuous variable in **X** leads to usage of inexact matching estimators, such as distance-based measures as in Abadie and Imbens (2011) and propensity score approaches as first employed by Rosenbaum and Rubin (1983). See Caliendo and Kopeinig (2008) for a thorough overview of the implementation of propensity score matching. These approaches produce a measure of the ATT based upon

$$\gamma^{ATT} = E(Y_1|g(X), D=1) - E(Y_0|g(X), D=0), \quad (6)$$

differing only in the construction of g(X) but all assuming implicitly or explicitly that $(Y_0 \perp D)|g(X)$, i.e., that conditional on **X** treatment is as good as randomly assigned. This is the so-called conditional independence assumption.

In Abadie and Imbens (2006, 2011) a measure of distance between households (e.g., the Euclidean or Mahalanobis distance) is constructed,[6] and the closest match to a treated household is thus the control household at the shortest distance from the treated household. Abadie and Imbens (2011) show that in this case, the matching estimator in Eq. (5) is biased for the true ATT, propose a regression-based bias correction, and derive the asymptotic variance of the bias-corrected estimator, which is asymptotically normal (Abadie and Imbens, 2006).

A convenient and computationally less intensive alternative is to deploy propensity score matching, which relies on the fact that conditioning on the propensity score (a single-index value) is equivalent to conditioning on **X**. One first fits a logit or probit model to explain treatment status as a function of the covariates **X**, and computes a predicted probability of treatment p_i for each household. Each treated household is matched with the control household(s) with the closest p_i, and the ATT is computed using Eq. (6) under a variety of weighting schemes for $w(i,j)$ from Eq. (5).[7] One then checks that the covariates are balanced post matching, which hopefully implies that conditional independence is satisfied. However, neither approach guarantees that the matched samples will be balanced with respect to the covariates **X**. Both approaches can be relatively time-consuming to implement. Iacus et al. (2011) propose coarsened exact matching (CEM) to get around these two limitations.

Table 4
Coarsened exact matching (CEM) results. QHEC households v. control households.

	CEM 1	CEM 2	CEM 2
Matching variables	–2008 usage	–House characteristics	–2008 usage –House characteristics
Matched households	10,580 (97.83%)	8770 (81.09%)	3603 (33.33%)
Nobs available for the final regression	188,130 (97.91%)	156,043 (81.21%)	64,366 (33.50%)

With CEM, continuous variables are converted to discrete interval data, and exact matching strata are constructed. The algorithm that implements this conversion seeks to select intervals that make the treated units and their matches among the controls balanced with respect to **X**. The procedure produces weights. Unmatched units receive a weight of zero. Matched units receive a weight equal to one if they belong to the treatment group, and $\frac{m_C}{m_T}\cdot\frac{m_T^s}{m_C^s}$ if they belong to the control group, where m_C is the total number of control units, m_T is the total number of treatment units, and m_C^s and m_T^s are their counterparts in stratum s. The weights make the treatment and control groups balanced with respect to **X**.[8]

Finally, one runs regression (3), where the right-hand side is augmented with the matching variables to control for any residual imbalance, by weighted least squares, where the weights are the CEM weights. Iacus et al. (2011) compare various matching approaches using Monte Carlo simulations and conclude that CEM outperforms the others in terms of bias and variance of the ATT, as well as execution time. For this reason, we deploy CEM in this paper as our primary matching method, and run the final weighted least square regression using the full panel dataset. Ferraro and Miranda (2014) employ a similar approach.

We perform each matching exercise three times, first using energy usage in the winter and the summer of 2008 (well before participation in the utility's programs) as the matching variables, then with house characteristics (including the type of heating system) as the matching variables, and then again with the broader set of matching variables—namely 2008 winter and summer usage *and* house characteristics. The first approach considers a treatment and control household a matched pair if their 2008 winter and summer electricity consumptions levels were roughly the same. We expect the third approach to be the most stringent: the two households would not be considered good matches for each other if, for example, one of them had a very large house and the other a very small house, as the implied energy intensities would be very different.

We wish to check if the estimation results are very sensitive to using a coarser matching criterion (prior usage only), which presumably yields more matched households, versus a more stringent one, which is expected to yield fewer matches, for a smaller final sample size. We emphasize that the former approach is easily deployed when the only information about households available to the researcher is their usage itself (i.e., the billing and usage data from the utility), while the latter is possible only when usage data are merged with household or house structure information.

As a robustness check, we also estimate the ATT using both a traditional matching method based on minimizing the Mahalanobis distance and propensity score matching. Both use cross-sectional samples from

[6] Consider the vector of covariates **X** for treated unit i, and let $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ denote the vector of means and the variance–covariance matrix of the covariates in the control unit. The Mahalanobis distance is defined as the square root of $(\mathbf{X}-\boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\mathbf{X}-\boldsymbol{\mu})$.

[7] In practice, by changing the definition of $w(i,j)$ it is possible to identify multiple matches for each treated household including kernel approaches that weight "near" observations more heavily than distant observations or uniform approaches such as single or many nearest neighbors each weighted equally. They may also impose additional requirement on the matches (for instance, that they lie within a specified radius or "caliper" around each treated unit).

[8] Unless restrictions are imposed on the CEM algorithm by the researcher, CEM will by default uses all possible matches for the treated units, and is thus different from distance-based approaches or propensity score matching, where the number of matches used to estimate the ATT is arbitrarily defined by the researcher. Using CEM allows more matches where the counterfactuals are thick and fewer matches where good counterfactuals do not exist. While sharing these characteristics with kernel matching algorithms the CEM in practice allows the bandwidth to vary per discretized observation where the kernel imposes a global bandwidth.

Table 5
Weighted means of matching variables in the QHEC treatment and control groups.

Model	Variable	Control group	QHEC treatment group
CEM1	Summer 2008 usage	5779.081	5745.573
	Winter 2008 usage	6572.229	6618.136
CEM 2	Summer 2008 usage	5933.610	5804.570
	Winter 2008 usage	6512.563	6760.096
	Square feet	1095.945	1912.586
	Heat pump	0.61096	0.61096
	Vintage	3.912329	3.912329
	SF home	0.715068	0.715068
	Brick	0.101370	0.101370
	Frame	0.073972	0.073972
	Siding	0.745206	0.745206
CEM3	Summer 2008 usage	5424.801	5404.673
	Winter 2008 usage	6178.023	6206.095
	Square feet	1788.557	1788.143
	Heat pump	0.642857	0.642857
	Vintage	4.030612	4.030612
	SF home	0.70068	0.70068
	Brick	0.061225	0.061225
	Frame	0.054422	0.054422
	Siding	0.836735	0.836735

control and program participating households from season s and year t, where s and t are post-treatment periods for the participating households. For consistency with Eq. (3), the outcome variable is the difference between log electricity usage in season s in year and its counterpart in the same season the prior year.

5. Results

5.1. Main results

We begin our discussion of the estimation results with those for the QHEC energy audit treatment. A summary of the CEM procedure where the matching variables are 2008 winter and summer usage is reported in Table 4. We dub this "CEM 1." Most of the households, and their seasonal usage totals, are retained in the final regressions. The CEM 1-weighted averages of the matching variables (the household's winter and summer usage in 2008) are virtually identical across the treatment and control households (first panel of Table 5).

Table 6, column (A), reports the results from fitting the triple difference model without attempting to trim the sample or attain covariate balance. The QHEC appears to reduce usage by 2.74%, but this effect is only marginally statistically significant at the 5% level. When the same model is re-run with the CEM 1 weights, the average treatment effect of participating in the energy audit program is similar, and statistically weaker (the t statistic is −1.75, which indicates significance at the 10% level). When attention is restricted to summertime billing cycles (columns (C) and (D)), participation in the home energy audit program brings a slightly stronger reduction in energy use (a 3.3% decline), which is again marginally statistically significant at the 5% level.

In the second CEM procedure ("CEM 2") house characteristics are the matching variables. This variant is able to match 81% of the homes and retains some 81% of the observations in the final regression (see Table 4). Despite the smaller sample size, the regression results (shown in Table 7) indicate that the ATT is a 3–4% reduction in usage (depending on the specification) The effect is statistically significant at the 1% or better, and the t statistics larger in absolute value than those with CEM 1. The ATT is, again, slightly stronger in the summer.

The third CEM approach ("CEM 3") uses 2008 winter and summer usage *and* dwelling characteristics to create the matching strata. As shown in Table 4, this discards many more observations than CEM 1. As before, good balance is achieved in terms of the covariates across treated and control units. Only about one-third of the available observations are retained in the final regression. The CEM 3-weighted averages of the matching variables are, again, practically identical across the treatment and control groups (second panel of Table 5).

In spite of the dramatically smaller sample used in the final regression (where we control for 2008 electricity usage *and* dwelling characteristics), when applying the CEM 3 weights the average treatment effect of QHEC is stronger, indicating a decline in usage by up to 5.5%. Summertime savings in electricity usage are of similar magnitude, and likewise statistically significant at the conventional levels (Table 8).

Turning attention to the other treatment—the heat pump rebate—the results of the CEM algorithms are similar to those with the home energy audit. If the matching variables are limited to 2008 winter and summer usage levels, then some 98% of the households are matched. If house characteristics are the matching variables, about 90% of the homes are matched and 90% of the observations retained in the estimation sample. When past usage and house characteristics are used, only about 46% of the homes are matched, and so the final sample size for the regression is greatly reduced (Table 9).

We run the triple difference model of Eq. (1) without any weights or trimming the sample, and the results are displayed in the first column of Table 10. They indicate that participating in the heat pump rebate program (which means that the existing heat pump is replaced with an energy-efficient one) brings a 5.3% reduction in energy usage. On trimming the sample and applying the CEM weights, the average treatment effect of changing the heat pump ranges between a 4 and 5% reduction in electricity usage, and remains strongly statistically significant (Tables 10–12).

Quantifying the summertime savings is, however, more difficult. If the CEM 1 or CEM 2 weights are applied, replacing the heat pump seems to produce 3.5–3.7% reductions in electricity usage, but the summertime average treatment effect is much smaller (about 2%), and statistically insignificant, when we use the CEM 3 weights (Table 12).

Table 6
DDD model. Dep. variable: ln electricity use. Treatment: quick home energy audit. T statistics in parentheses.

	Full DDD No weights	Full DDD CEM 1 weights	Simplified DDD (summers 2010–11) CEM 1 weights	Simplified DDD (summers only) CEM 1 weights
Regressors or fixed effects from Eq. (1)				
Treatment dummy	−0.0278	−0.0250	−0.0340	−0.0344
	(−2.04)	(−1.75)	(−2.15)	(−2.26)
Household × season FE	Yes	Yes	Yes	Yes
Season × year FE	Yes	Yes	n/a	Yes
Household × year FE	Yes	Yes	No	No
Weather controls	Yes	Yes	Yes	Yes
Matching variables to control for any residual imbalance				
Benchmark year usage	No*	No*	Yes	Yes
House characteristics	No	No	No	No
Number observations	108,387	106,501	16,125	26,394

* Absorbed into the household-year FE.

Table 7
DDD model. Dep. variable: ln electricity use. Treatment: quick home energy audit. T statistics in parentheses.

	Full DDD no weights	Full DDD CEM 2 weights	Simplified DDD (summers 2010–11) CEM 2 weights	Simplified DDD (summers only) CEM 2 weights
Regressors or fixed effects from Eq. (1)				
Treatment dummy	−0.0278	−0.0316	−0.0380	−0.0388
	(−2.04)	(−2.29)	(−4.57)	(−2.66)
Household × season FE	Yes	Yes	Yes	Yes
Season × year FE	Yes	Yes	Yes	Yes
Household × year FE	Yes	Yes	No	No
Weather controls	Yes	Yes	Yes	Yes
Matching variables to control for any residual imbalance				
Benchmark year usage	No	No	No	No
House characteristics	No*	No*	Yes	Yes
Number observations	108,387	90,529	13,778	22,530

* Absorbed into the household-year FE.

Table 8
DDD model. Dep. variable: ln electricity use. Treatment: quick home energy audit. T statistics in parentheses.

	Full DDD no weights	Full DDD CEM 3 weights	Simplified DDD (summers 2010–11) CEM 3 weights	Simplified DDD (summers only) CEM 3 weights
Regressors or fixed effects from Eq. (1)				
Treatment dummy	−0.0278	−0.0544	−0.0480	−0.0479
	(−2.04)	(−3.63)	(−4.61)	(−2.86)
Household × season FE	Yes	Yes	Yes	Yes
Season × year FE	Yes	Yes	Yes	Yes
Household × year FE	Yes	Yes	No	No
Weather controls	Yes	Yes	Yes	Yes
Matching variables to control for any residual imbalance				
Benchmark year usage	No*	No*	Yes	Yes
House characteristics	No*	No*	Yes	Yes
Number observations	108,387	37,511	5780	9338

* Absorbed into the household-year FE.

5.2. Robustness checks

The results discussed so far are based on utilizing the full panel dataset—obviously our preferred approach. We also created cross-sections from the existing panel dataset, and used them to apply propensity score and distance-based matching algorithms. The results derived from these estimation approaches provide useful robustness checks. Tables A.1 and A.2 in the Appendix present results from the PSM and distance based estimators across multiple choices for nearest neighbor, again with and without structural characteristics included in the matching variables.

There is ample evidence of variability in these results if one compares results across matching approaches or within approaches using different numbers of neighbors or a richer set of matching variables (usage and dwelling data). In both rebate and audit treatments the strongest effects arise from the heaviest usage period during the summer months with the audit reduction of 4 to 5% and the rebate reduction in differences of ~3%. While the preponderance of results are of similar magnitude and significance to our preferred specification and approach (CEM plus the panel), this variability and reliance on the researcher's choice of the number of neighbors make the CEM weighting approach all the more attractive.

6. Conclusions

We have used a unique set of data from Maryland that combines electricity usage levels and utility program participation records with structural characteristics of the dwelling to estimate the electricity usage reductions that can be attributed to residential energy audits and incentives to replace existing heat pumps with new, and more energy efficient, ones. We have observations on usage for participating households and for a group of similar, non-participating households, which we regard as control units. We have observations on usage before and after program participation (which took place in Q1 2011) for all households.

Since program participation is voluntary, naïve estimates of its effects are likely affected by selection bias, which we have attempted to address by deploying household-by-season fixed effects, season-by-year fixed effects, and household-by-year fixed effects, plus matching methods to restore a quasi-experiment design. Most applications of matching methods in economics are for cross-sections. By contrast, our dataset is a panel, and we fully exploit it by applying coarsened exact matching on *households* and *then* running regressions that use the full panel of observations on usage.

Our findings suggests that past usage alone—as is often done in studies that lack information among other determinants of residential

Table 9
Coarsened exact matching (CEM) results. Heat pump households v. control households with heat pumps.

	CEM 1	CEM 2	CEM 3
Matching variables	–2008 usage	–House characteristics	–2008 usage –House characteristics
Matched households	6681 (97.45%)	6240 (91.01%)	3173 (46.28%)
Nobs available for the final regression	112,326 (97.65%)	110,018 (90.27%)	56,797 (46.60%)

Table 10
DDD model. Dep. variable: ln electricity use. Treatment: heat pump rebate. T statistics in parentheses.

	Full DDD No weights	Full DDD CEM 1 weights	Simplified DDD (summers 2010–11) CEM 1 weights	Simplified DDD (summers only) CEM 1 weights
Regressors or fixed effects from Eq. (1)				
Treatment dummy	−0.0546	−0.0430	−0.0375	−0.0373
	(−3.15)	(−2.68)	(−3.01)	(−3.15)
Household × season FE	Yes	Yes	Yes	Yes
Season × year	Yes	Yes		Yes
Household × year FE	Yes	Yes	No	No
Weather controls	Yes	Yes	Yes	Yes
Matching variables to control for any residual imbalance				
Benchmark year usage	No*	No*	Yes	Yes
House characteristics	No	No	No	No
Number observations	68,754	67,604	10,209	16,735

* Absorbed into the household-year FE.

Table 11
DDD model. Dep. variable: ln electricity use. Treatment: heat pump rebate. T statistics in parentheses.

	Full DDD No weights	Full DDD CEM 2 weights	Simplified DDD (summers 2010–11) CEM 2 weights	Simplified DDD (summers only) CEM 2 weights
Regressors or fixed effects from Eq. (1)				
Treatment dummy	−0.0546	−0.0595	−0.0350	−0.0355
	(−3.15)	(−3.47)	(−2.67)	(−2.82)
Household × season FE	Yes	Yes	Yes	Yes
Season × year	Yes	Yes	No	Yes
Household × year FE	Yes	Yes	No	No
Weather controls	Yes	Yes	Yes	Yes
Matching variables to control for residual imbalance				
Benchmark year usage	No	No	No	No
House characteristics	No*	No*	Yes	Yes
Number observations	68,754	64,256	9471	15,519

* Absorbed into the Household-year FE.

Table 12
DDD model. Dep. variable: ln electricity use. Treatment: heat pump rebate. T statistics in parentheses.

	Full DDD No weights	Full DDD CEM 3 weights	Simplified DDD (summers 2010–11) CEM 3 weights	Simplified DDD (summers only) CEM 3 weights
Regressors or fixed effects from Eq. (1)				
Treatment dummy	−0.0546	−0.0419	−0.0202	−0.0202
	(−3.15)	(−2.49)	(−1.52)	(−1.65)
Household × season FE	Yes	Yes	Yes	Yes
Season × year	Yes	Yes	No	Yes
Household × year FE	Yes	Yes	No	No
Weather controls	Yes	Yes	Yes	Yes
Matching variables to control for residual imbalance				
Benchmark year usage	No*	No*	Yes	Yes
House characteristics	No*	No*	Yes	Yes
Number observations	68,754	33,268	10,209	16,735

* Absorbed into the Household-year FE.

energy usage, such as house and household characteristics (e.g., Ito, 2014)—may not be not sufficient and that house characteristics are important. We find that residential energy audits reduce usage by about 5%, and that the heat pump rebate has an effect of similar magnitude.[9] These figures may include rebound effects, i.e., the increase in energy use thought to occur when improved energy efficiency lowers the price of a unit of energy services (Alberini et al., 2014).

For policy purposes, it is of interest to compute the cost-effectiveness of these programs, namely the cost per ton of CO2 emissions removed. With our programs, however, these calculations are not simple. Consider for example the QHEC program. Starting from a baseline of 18,000 kWh per year, a 5% reduction implies that 900 kWh are saved per year. Since we do not know whether these savings were attained with simple behavioral changes or by replacing equipment or making other energy-efficiency investments, it is difficult to say what the time horizon over which these savings are accrued is.

If we assume that it is 7 years (as assumed by the utility), then a participating household would avoid 3.830 t of CO2. Assuming that the cost of the audit to the utility is $200, and that an additional $60 worth of products are offered to the household, for a total of $260 per QHEC, then the cost per ton of CO2 emissions abated is $67.88.[10] The cost

[9] This is a greater saving that than estimated by the utility, which is 375 kWh a year. See http://www.smeco.coop/saveEnergy/quickHomeEnergyCheckup/comparisonChart.aspx (accessed 13 June 2014).

[10] We assume 0.608 kg of CO2 emissions per kWh generated.

falls to $47.50 per ton of CO emissions if we assume that the usage reductions would be sustained for 10 years. This is above the $21 "typical" social cost of carbon used by federal agencies in benefit–cost analyses, but well within the range of values in Greenstone (2013), which are obtained under various scenarios and discount rate assumptions.

With the heat pumps rebate, we assume that the lifetime of a heat pump is 10 years, a figure commonly indicated in utility and federal government agency calculations. This means that, starting from a baseline of 21,000 kWh a year and assuming a rebate of $400, the cost per ton of CO_2 emissions avoided is about $59. The problem with this calculation is that evidence from other studies (Alberini et al., 2014; Boomhower and Davis, 2014), and the high levels of usage observed in our own sample prior to replacing the heat pump, suggest that people replace heat pumps when their existing equipment is about to die and essentially free ride on the incentives. Since we find that replacing heat pumps with new and more efficient ones does indeed decrease energy usage, energy efficiency standards for new heat pumps might be sufficient to ensure such usage reductions, which would presumably occur at no additional cost to the entity issuing the rebates.

Appendix A

Table A.1
Summary of ATT estimates using Mahalanobis-distance and propensity score matching. QHEC.

		Distance based matching		Propensity score matching	
Number nearest neighbors	Outcome diff 2011 and 2010 by quarter	Usage and dwelling data	Usage data only	Usage and dwelling data	Usage data only
3	Q2	−0.036	−0.020	−0.030	−0.033
3	Q3	−0.051*	−0.037*	−0.048**	−0.064**
3	Q4	−0.014	−0.003	−0.011	−0.029**
5	Q2	−0.037	−0.019	−0.026	−0.033*
5	Q3	−0.051*	−0.036*	−0.046*	−0.048*
5	Q4	−0.011	−0.005	−0.011	−0.016
7	Q2	−0.038	−0.017	−0.023	−0.032
7	Q3	−0.053*	−0.038*	−0.044*	−0.053**
7	Q4	−0.012	−0.004	−0.013	−0.017
11	Q2	−0.035	−0.018	−0.025	−0.025
11	Q3	−0.054**	−0.036*	−0.040*	−0.050**
11	Q4	−0.013	−0.006	−0.014	−0.012

*** at 1%, ** at 5%, * at 10%.
Note: the standard errors for the distance-based matching estimation are based on the exact formulae in Abadie and Imbens (2011). The PSM standard errors are bootstrapped with 1000 replications.

Table A.2
Summary of ATT estimates using Mahalanobis-distance and propensity score matching. Heat pump rebate.

		Distance based matching—Mahalanobis		Propensity score matching	
Number nearest neighbors	Outcome diff 2011 and 2010 by quarter	Usage and dwelling data	Usage data only	Usage and dwelling data	Usage data only
3	Q2	−0.005	−0.004	−0.002	0.002
3	Q3	−0.025	−0.032**	−0.030	−0.030**
3	Q4	−0.022*	−0.030**	−0.004	−0.009
5	Q2	−0.006	−0.005	−0.006	−0.001
5	Q3	−0.024	−0.029**	−0.039**	−0.024
5	Q4	−0.020	−0.025*	−0.011	−0.003
7	Q2	−0.005	−0.004	0.000	0.000
7	Q3	−0.020	−0.029*	−0.035**	−0.016
7	Q4	−0.020	−0.021	−0.007	−0.004
11	Q2	−0.002	−0.004	0.001	0.006
11	Q3	−0.016	−0.028*	−0.032*	−0.024
11	Q4	−0.021	−0.017	−0.006	−0.003

*** at 1%, ** at 5%, * at 10%.
Note: the standard errors for the distance-based matching estimation are based on the exact formulae in Abadie and Imbens (2011). The PSM standard errors are bootstrapped with 1000 replications.

References

Abadie, Alberto, Imbens, Guido W., 2006. Large-sample properties of matching estimators for average treatment effects. Econometrica 74 (1), 235–267.

Abadie, Alberto, Imbens, Guido W., 2011. Bias-corrected matching estimators for average treatment effects. J. Bus. Econ. Stat. 29 (1), 1–11.

Alberini, Anna, Gans, Will, Velez-Lopez, Daniel, 2011. Residential Consumption of Gas and Electricity in the U.S The Role of Prices and Income. Energy Economics 33 (5), 870–881.

Alberini, Anna, Towe, Charles, Gans, Will, 2014. Energy efficiency incentives: do they work for heating and cooling equipment? Evidence from Maryland homeowners. The Energy Journal (forthcoming).

Allaire, Maura, Brown, Stephen, 2012. US Energy Subsidies: Effects on Energy Markets and Carbon Emission, Prepared for the Pew Charitable Trusts. http://www.pewtrusts.org/uploadedFiles/wwwpewtrustsorg/Reports/Fiscal_and_Budget_Policy/EnergySubsidiesFINAL.pdf.

Angrist, Joshua D., Pischke, Jörn-Steffen, 2009. Mostly Harmless Econometrics. An Empiricist's Companion, Princeton. Princeton University Press, NJ.

Auffhammer, Maximilian, 2014. "Quantifying Intensive and Extensive Margin Adaptation Responses to Climate Change: A Study of California's Residential Electricity Consumption", Work in Progress. University of California, Berkeley.

Auffhammer, Maximilian, Blumstein, C., Fowlie, Meredith, 2008. Demand side management and energy efficiency revisited. Energy J. 29 (3), 91–104.

Boomhower, Judon, Davis, Lucas, 2014. Credible Approach for Measuring Inframarginal Participation in Energy Efficiency Programs, Journal of Public Economics. http://www.sciencedirect.com/science/article/pii/S0047272714000589.
Caliendo, Marcus, Kopeinig, Sabine, 2008. Some practical guidance for the implementation of propensity score matching. J. Econ. Surv. 22 (1), 31–72.
Davis, Lucas, 2008. Durable goods and residential demand for energy and water: evidence from a field trial. RAND J. Econ. 39 (2), 530–546.
Dehejia, Rajeev, Wahba, Sadek, 1999. Causal effects in nonexperimental studies: reevaluating the evaluation of training programs. J. Am. Stat. Assoc. 94, 1053–1062.
Dehejia, Rajeev, Wahba, Sadek, 2002. Propensity score matching methods for nonexperimental causal studies. Rev. Econ. Stat. 84 (1), 151–161.
Dinan, Terry M., 2013. Federal Financial Support for Fuels and Energy Technologies, Testimony Before the Subcommittee on Energy, Committee on Science, Space and Technologies. US House of Representatives, Washington, DC (March, available at http://www.cbo.gov/publication/43993,).
Faruqui, Ahmad, Sergici, Sanem, Sharif, Ahmed, 2010. The impact of informational feedback on energy consumption. A survey of the experimental evidence. Energy 35, 1598–1608.
Ferraro, Paul, Miranda, Juan Jose, 2014. Panel Data Designs and Estimators as Alternatives for Randomized Controlled Trials in the Evaluation of Social Programs available at, http://www2.gsu.edu/~wwwcec/docs/Ferraro%20and%20Miranda%20Panel%20Data%20Rep%20POST.pdf (last accessed 26 April 2015).
Gans, Will, Alberini, Anna, Longo, Alberto, 2013. Smart meter devices and the effect of feedback on residential electricity consumption: evidence from a natural experiment in Northern Ireland. Energy Econ. 36, 729–743.
Greenstone, Michael, Kopits, Elizabeth, Wolverton, Ann, 2013. Developing a social cost of carbon for US regulatory analysis: a methodology and interpretation. Rev. Environ. Econ. Policy 7 (1), 23–46.
Hartman, Raymond S., 1988. Self-selection bias in the evaluation of voluntary energy conservation programs. Rev. Econ. Stat. 70 (3).
Hassett, Kevin A., Metcalf, Gilbert E., 1995. Energy tax credits and residential conservation investment: evidence from panel data. J. Public Econ. 57, 201–217.
Iacus, Stefano M., King, Gary, Porro, Giuseppe, 2011. Causal inference without balance checkin: coarsened exact matching. Political Analysis (COMPLETE REF).
Intergovernmental Panel on Climate Change, 2007. Fourth Assessment Report available at http://www.ipcc.ch/publications_and_data/publications_and_data_reports.shtml#1.
Ito, Koichiro, 2014. Do consumers respond to marginal or average price? Evidence from nonlinear electricity pricing. Am. Econ. Rev. 104 (2), 537–563.
Jessoe, Katrina, Rapson, David, 2014. Knowledge is (less) power: experimental evidence from residential energy use. Am. Econ. Rev. 104 (4), 1417–1438.
Joskow, P.L., Marron, D.B., 1992. What does a negawatt really cost? Evidence from utility conservation programs. Energy J. 13, 41–73.
Loughran, D., Kulick, J., 2004. Demand side management and energy efficiency in the United States. Energy J. 25 (1), 19–41.
McKinsey and Company, 2009. Unlocking Energy Efficiency in the US Economy available at http://www.mckinsey.com/client_service/electric_power_and_natural_gas/latest_thinking/unlocking_energy_efficiency_in_the_us_economy.
Rosenbaum, Paul, Rubin, Donald, 1983. The central role of the propensity score in observational studies for causal effects. Biometrika 70 (1), 41–55.
Waldman, Donald M., Ozog, Michael T., 1996. Natural and incentive-induced conservation in voluntary energy management programs. South. Econ. J. 62 (4).