Smart Meter Data Predict Household Propensity to Enroll in Energy Efficiency Programs

Michael Zeifman, Fraunhofer Center for Sustainable Energy Systems, Boston MA

ABSTRACT

Opt-in energy efficiency and demand response programs require consumers to enroll. Conventionally, households are recruited to participate in these residential programs without regard to the consumers' propensity to enroll in such programs. Since the fraction of successfully recruited households is usually low, effectiveness of these recruitment efforts is questionable. Moreover, recruitment within a sub-population that is likely to enroll could be made more effective by targeted marketing.

The state-of-the-art methodology method for household enrollment prediction involves running measurable household characteristics (e.g., age, household income, education, presence of children, average energy bill) through a mathematical tool, (e.g., a multivariate logistic regression) that connects these predictor variables with the probability to enroll. The estimation of the regression coefficients (i.e., training) typically requires data on about a thousand households that enrolled and another thousand households that did not enroll in previous program(s). Unfortunately, the prediction accuracy of this method is just slightly better than 50%, and the required household data are not freely available to utilities/ program contractors.

This paper describes a new method to predict household propensity to enroll in opt-in behavioral programs. The only data needed to complete this analysis is 12 months' of hourly electricity consumption data from households' smart meters. The method uses advanced machine learning algorithms to reach an unprecedented prediction accuracy of about 90%. The results are based on our study of a West Coast behavior-based residential program.

Introduction

Behavior-based energy efficiency programs are designed to affect consumer behaviors to achieve energy and/or peak demand savings [DOE 2012]. Naturally, customers tend to self-select into the program; those that voluntary enroll in a behavior-based program are most likely to benefit from it [DOE 2012]. Examples of the opt-in programs are Wisconsin PowerCost Monitor in-home energy display [DOE 2012] and time-of-use residential electricity rate [DOE 2013] programs.

In the US, the enrollment rate of such opt-in programs, defined as the number of recruited customers divided by the number of solicited customers, is often a small fraction. For example, in a recent DOE study of time-based residential electricity rate programs, the opt-in enrollment rates ranged from 5% to 28% with the average enrollment rate being 11% [DOE 2013]. These numbers indicate there is potential of increasing the effectiveness of recruitment efforts..

Evaluation of behavioral-based programs usually involves a control group of households that do not receive treatment. Such a control group needs to be as similar to the treatment group as possible. Unfortunately, experimental design in which this similarity is preserved by random assigning of the recruited households into a treatment group and a control group (Randomized Controlled Trials - RCT) is not always possible and/or suitable.

One popular non-experimental method of constructing a control group is matching energy consumption (e.g., pre-treatment energy bills) and other available observable characteristics (e.g., income level or household square footage) of the enrolled households and the candidate nearby households that did not have a chance to enroll [Allcott 2011]. This method can create a non-experimental control group with the observable characteristics being very similar to those of the treatment group, yet it is known to generate the so-called selection bias in energy saving estimation [DOE 2012]. Allcott, for example, reports a discrepancy of up to 300% between program savings estimated using an experimental control group and a matched non-experimental control group [Allcott 2011].

When using this matching technique, the main difference between the non-experimental control and the experimental groups is that the households in the latter group exhibit desire to enroll in the program. The households in the former group may or may not have such enrollment propensity; in light of the generally low enrollment rates, it is actually unlikely that these households would have enrolled had they been given a chance to enroll (see section Illustrative Example for an example of actual enrollment rate). On the other hand, if we could reliably predict the propensity to enroll for a given household, we would be able to eliminate the selection bias in constructing the non-experimental control groups by selecting only households that are likely to enroll. Therefore, prediction of household propensity to enroll is important for both reduction of the recruitment costs and construction of valid non-experimental control groups.

Application of a propensity estimation method commonly used in social science [Caliendo & Kopeinig 2008] to predict enrollment in opt-in behavioral energy efficiency programs is not straightforward. This is because it is unclear which socio-economic variables can be used as predictors and what the underlying causation would be. To the best of our knowledge, the only published work in this field is a paper by Harding and Hsiaw on evaluation of a behavior based opt-in energy efficiency program in Northern Illinois with a non-experimental control group [Harding & Hsiaw 2014]. Harding and Hsiaw attempted to predict household enrollment using household data, enrollment results (enrolled or not) and logistic regression. Even though Harding and Hsiaw used potentially strong predictor variables such as "environmental issues" (i.e., magazine subscriptions and/or mail response indicating household's interest in environmental or wildlife issues) and "green living" (i.e., buying green household cleaning products, eating organic foods, donating funds to environmental causes, or driving hybrid vehicles), their goodness of fit as expressed by (pseudo) R-squared was mere 0.035[1]. Whereas a connection between the pseudo R-squared in logistic regression and the classification accuracy (households enrolled versus not enrolled) may not be straightforward, it is clear that the classification accuracy of this method is just slightly better than 50%.

This problem is further exacerbated by unavailability of socio-economic household data to utilities and/or program contractors as well as privacy concerns. Harding and Hsiaw [Harding & Hsiaw 2014] used socio-economic data that included, in addition to the mentioned "environmental issues" and "green living," traditional variables such as age, household income, presence of children, education, smoking and gambling habits, dieting, home loan-to-value ratio and presence of department store lines of credit. All these variables needed to be purchased [Harding & Hsiaw 2014].

In this paper, we describe a new method for prediction of household propensity to enroll in opt-in behavioral energy efficiency/demand response such programs. This method only requires hourly electricity consumption data from households' smart meters, collected over 6-12 months. With the smart meters being massively installed in US residences [Edison 2012], the household interval energy consumption data are freely available to some US utilities. Unlike socio-economic variables, household interval electricity consumption data cannot be easily used for household identification; thus, this method reduces privacy concerns. Finally, this method implements advanced machine learning

[1] Table 2 of [Harding & Hsiaw 2014] uses term "R-squared" for the logit model

algorithms to reach an unprecedented prediction accuracy of about 90%. Given the new application of this model we present the findings from one utility, but believe this will be applicable across the nation.

In the next sections, we give a background on the propensity score matching method of constructing non-experimental control groups and explain the method in detail. We also provide a case study of an evaluation of a behavior based energy efficiency program and comparative analysis of our method and a traditional non-experimental method. The paper concludes with a Summary where we discuss various aspects of the method including its implementation.

Propensity Score Estimation Using Hourly Energy Consumption

The propensity score matching (PSM) method constructs a non-experimental control group on the basis of the propensity score, i.e., the probability of participating in a program given observed characteristics, or covariates [Caliendo & Kopeinig 2008]. Once the probability of participation can be estimated, the control group is constructed to closely follow the participation probabilities calculated for the treatment group. The probability of participation can be substituted by the underlying index of the probability estimation for the propensity score as the latter quantity better differentiates between observations in the distribution extremes [Lechner 2000]. In this case, the propensity score is not bounded to the interval from 0 to 1.

As we show in the previous section, socio-economic household variables are difficult to use for propensity estimation, partly because they are not strong predictors of enrollment and partly because they are usually not freely available to utilities/ cause privacy issues. On the other hand, hourly household electricity consumption data, increasingly available to utilities that install residential smart meters [Edison 2012], embed significant information about household behaviors. Using electricity consumption data obtained at hourly resolution, for example, researchers can deduce the presence of major household electric appliances and their operational profiles by applying relatively simple disaggregation algorithms [Birt et al. 2012]. Therefore, we assume that the hourly electricity data of residential utility customers collected over a prolonged period of time prior to treatment (e.g., one year) can be a predictor of the propensity to enroll.

For the binary treatment, the usual choice of a model for propensity estimation is either a linear or a logit probability model [Caliendo & Kopeinig 2008]. Such models are feasible for traditional low-dimensional socio-econometric covariates, e.g., household income or house floorspace. The sheer size of the energy consumption data array of a household (8,760 data points for one year), however, makes direct application of the linear/logit models unfeasible.

The binary treatment problem underlying the propensity score is very close to a binary classification problem. Moreover, Heckman et al. [Heckman et al. 1998] used classification accuracy to assess goodness of fit of their logit model for propensity to enroll. Therefore, instead of using a probability model, we use a binary classification scheme to infer the propensity score.

This patent-pending method is explained in detail elsewhere [Zeifman 2014a], [Zeifman 2014b]. Briefly, a nonlinear machine learning (NML) algorithm maps a time series of a household hourly data onto a pair of numbers, or scores. The first score in the pair indicates the score for class one (enrolled households), while the second score indicates the score for class two (not enrolled households). For either class, we want the NML algorithm to yield the score as close to unity as possible in case of the data array belonging to this class, and as close to zero as possible otherwise. To this end, we use data from the enrolled and unenrolled households for algorithm training, i.e., fitting to the data. Figure 1 illustrates this principle. Once the algorithm is trained, it can be used for household classification: the class with the higher score wins.

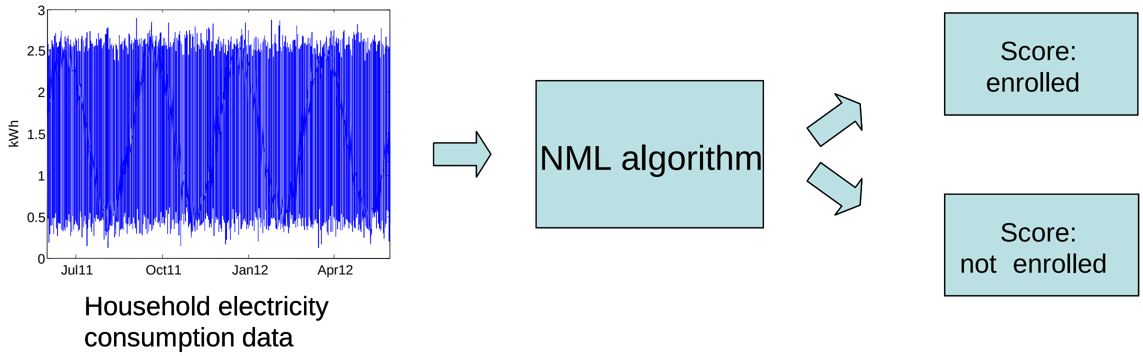

Figure 1. NML algorithm calculates two scores for hour-resolution electricity data from a household.

Note that term "training," which is usually used in the machine learning community, means estimation of the parameters of the underlying classification method using experimental data. In our case we estimate a matrix of weights that maps a 8,760x1 household electricity consumption data array onto the two numeric scores. In this sense, the conventional PSM methods also use training since the parameters need to be estimated using the socio-economic independent variables and the binary enrollment outcome as the dependent variable.

To illustrate how the method works, we consider a case study next.

Illustrative Example

In 2012, a major West Coast Utility Company partnered with a private contractor to launch a new opt-in behavior-based residential energy efficiency program[2]. Participants were recruited by several channels, including local educational institutes, social media, and news advertisement. The participants could control their electricity usage by monitoring their hourly electricity consumption data. Significant awards were offered for energy savings to the participating households.

To be eligible for this Program, residential customers must reside within a specific geographic area (section of a major city). Out of approximately 470,000 eligible customers, about 5,600 customers enrolled between June 2012 and September 2012.

We received information on about 5,600 households that enrolled in the Program and on about 32,000 households located outside the Program area but still within the same city and micro climate zone(s). This information included:

- Household electricity consumption with hourly resolution for a minimum of twelve months before the program started (i.e., on June 1, 2011 or earlier);
- Household zip code.

Due to strict utility regulations, no personal information accompanied the household data.

There are no statistically significant differences between the two household populations in terms of characteristics available to us. The average hourly energy consumption of the 5,600 enrolled households during the pre-program period was 0.5957 kWh and the standard deviation of this average was 0.012 kWh. The average hourly energy consumption of the 32,000 not-enrolled households during the pre-program period was 0.6126 kWh and the standard deviation of the average was 0.007 kWh. Therefore, the difference between the average energy consumptions of these two large samples is within two standard deviations, i.e., statistically insignificant.

[2] Our research agreement prevents us from publishing the contractor name. The contractor requested us to keep the utility name confidential.

Also, we assessed whether any observed differences could be attributed to differences in socio-economic population parameters. For example, if the enrolled households were located in a wealthy neighborhood and the not-enrolled households were located in a poor neighborhood, the latent differences in energy consumption could be attributed to the differences in those socio-economic population parameters. We used the US Census data to assess potential differences within and between the enrolled and not-enrolled populations. We studied parameters such as median age, fraction of family households, average family size, and median household income for the zip codes related to the two populations. We found no statistically significant difference between these parameters in the two populations.

Training and testing. In the mentioned papers on propensity score matching [Heckman et al. 1998], [Harding & Hsiaw 2014], the goodness of prediction was evaluated by the same experimental data that were used in fitting the model. In other words, the same data were used for both training and testing.

Using the same data sets for training and testing may lead to the overfitting problem [Dietterich 2000]. The goodness of fit would be unrealistically high for the data sets used but could be much lower for new data. Accordingly, we decided to use different samples for training and testing of our NML algorithm. To this end, we selected at random a sample of 2,000 enrolled households (out of 5,600 total) and a sample of 2,000 not enrolled households[3] (out of 32,000 total) for algorithm training. To test the classification performance, we selected at random a different sample of 2,000 enrolled households (out of 5,600 – 2,000 = 3,600 total) and a different sample of the not enrolled households (out of 32,000 – 2,000 = 30,000). Each of the four samples comprised 2,000 data arrays, each array being pre-treatment hourly electricity consumption of a household (i.e., 8,760 data points collected by Utility from June 1, 2011 to May 31, 2012).

As discussed earlier, and illustrated in Figure 1, in training, we need to estimate parameters of the NML algorithm (matrix of weights – [Zeifman 2014a]) by fitting simultaneously:

- the enrollment score - to be as close to unity as possible for the enrolled households
- the enrollment score - to be as close to zero as possible for the not-enrolled households
- the not-enrollment score - to be as close to zero as possible for the enrolled households
- the not-enrollment score - to be as close to unity as possible for the not-enrolled households.

In testing, we apply the trained algorithm to new data to calculate how many households were classified correctly. To provide fair comparison to the conventional PSM method, we also used for testing the samples that were used for training. The results are listed in Table 1.

As we can see in the table, our NML algorithm actually performs better than the logistic regression in the social PSM application: Heckman et al. report 81.89% of enrollment prediction accuracy and 81.96% of not-enrollment prediction accuracy using the same pair of samples for training and testing [Heckman et al. 1998]. The performance of NML algorithm slightly deteriorates if different samples are used for training and testing, but the accuracy still remains above 90%. One interesting observation in Table 1 is that the fraction of enrolled households classified as enrolled is about 9%, which is very close to the average enrollment rate of 11% reported in [DOE 2013]. Does this mean that the enrollment rate in the particular opt-in behavior-based energy efficiency program considered could be higher if better advertisement were used?

[3] Actually, these households are not-treatment households as they did not get a chance to enroll but given the very low enrollment rate (1.2%), we can safely use these households as those not enrolled. We hope that the reader, unlike us, will be given data of the eligible not enrolled households.

Table 1. Classification results of NML algorithm based on hourly household electricity consumption data from the pre-treatment period.

	Random sample of enrolled households, used for training (n = 2,000)	**Different random sample of enrolled households, used for testing** (n = 2,000)	**Random sample of not-enrolled households, used for training** (n = 2,000)	**Different random sample of not-enrolled households, used for testing** (n = 2,000)
Classified as enrolled	1,848 (92.4%)	1,825 (91.2%)	166 (8.3%)	191 (9.6%)
Classified as not enrolled	152 (7.6%)	175 (8.8%)	1,834 (91.7%)	1,809 (90.4%)

To further test the prediction accuracy of our method, we performed multiple cross validations using random sub-sampling. In this setting, we repeated the above-described process 1,000 times, drawing at random non-overlapping samples of 2,000 households for training and testing [Zeifman 2014a]. The results demonstrated accuracy levels exceeding 90% with 95% confidence intervals ranging from ±1.1% to ±1.4 %.

Distribution of propensity score. Unlike the conventional PSM methods, our approach yields two scores for each household data set: score 1 for enrollment and score 2 for not enrollment. The enrollment score is related to the enrollment probability, and the not-enrollment score is related to the probability to not enroll. If the suggested scores are good metrics of tendency to enroll/not enroll, their corresponding distributions for the training and testing datasets should not differ statistically.

Figure 2 shows empirical cumulative distribution functions (CDFs) calculated for the two scores and two non-overlapping samples from the case study: a sample of enrolled households used for training and a sample of enrolled households used for testing. It is seen in the Figure that the distribution functions of the two samples practically coincide for each of the score. Quantitatively, a Kolmogorov-Smirnov test can be applied to test a statistical hypothesis that the distribution functions are equal. We have conducted this test using the empirical distribution functions; the result is that the null hypothesis is not rejected with p-values of 0.41 and 0.43 for the enrollment and not enrollment scores, respectively. Similar results were obtained for samples of not enrolled households.

Overlap and common support. The common support or overlap condition is an important requirement to a propensity score model [Caliendo & Kopeinig, 2008]. It ensures that persons (or households in our case) with the same set of covariate values have a positive probability of being both participants and non-participants. In fact, the selection bias, observed by Heckman et al. [Heckman et al. 1998] for a propensity-matched non experimental control group, can be attributed to the lack of overlap between the distributions of enrollment probability of the enrolled and not enrolled samples [Lechner 2000].To follow this reasoning, we need to check overlap of the distributions of the enrollment score for the enrolled and not enrolled samples. Also, since we use the difference between the enrollment and not enrollment scores to classify a household, we also need to check overlap between the corresponding distributions of the score differences.

Figure 3 shows histograms of these quantities. Unlike the enrollment probabilities, our scores are not bounded in the interval from zero to unity, consequently, the distributions have typical unimodal shape [Lechner 2000] (The distribution of participation probability by Heckman et al. was bimodal with peaks near the distribution extremes, i.e., near zero and unity [Heckman et al. 1998]). The overlap is clearly seen for both the enrollment score and the difference between enrollment and not enrollment scores.

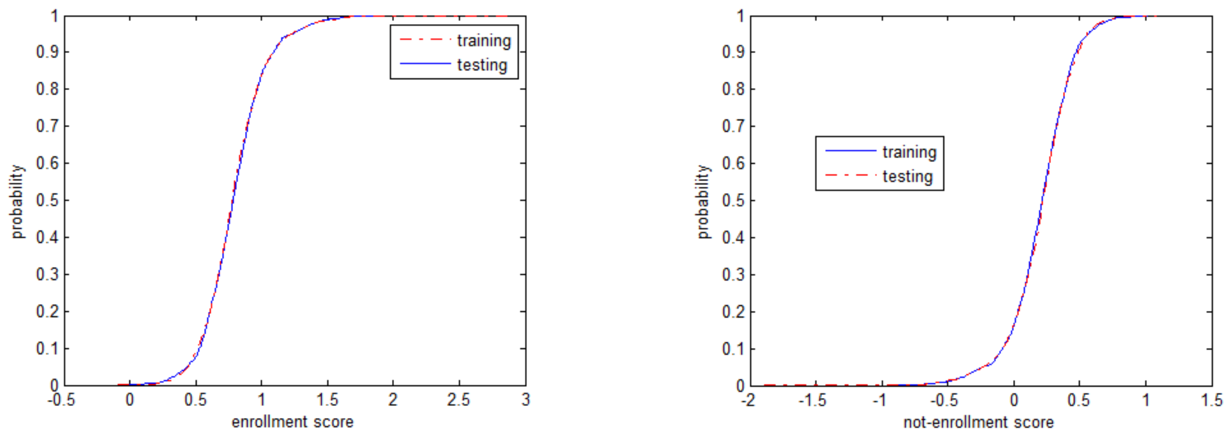

Figure 2. Empirical cumulative distribution functions of propensity score for enrollment (left) and not enrollment (right) of two different random samples (training and testing) of enrolled households. Size of each group is 2,000 households.

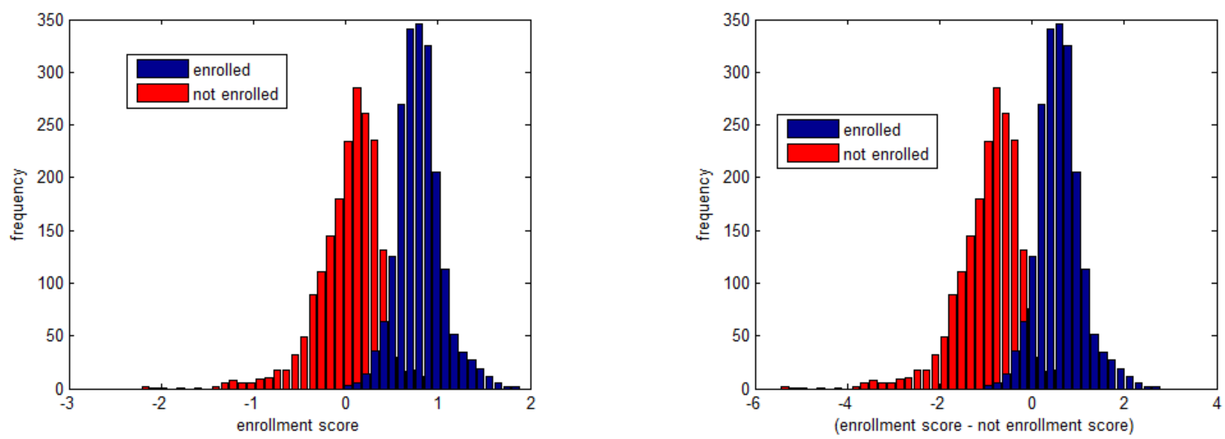

Figure 3. Enrollment score (left) and difference between enrollment and not enrollment scores for a sample of 2,000 enrolled households and a sample of 2,000 of not enrolled households

Construction of Non-Experimental Control Groups and Evaluation of Program Effect

We used the developed methodology to construct a non-experimental control group by PSM. For the sake of comparison, we also built a "conventional" non-experimental control group by matching household electricity consumption.

PSM Control Group Construction

Once the propensity score model has been fitted to experimental data, the next step in building a PSM control group is to calculate propensity scores of the enrolled and not-enrolled households. After the scores are available, an established matching method, e.g., the nearest neighbor matching or the stratification and interval sampling [Caliendo & Kopeing 2008] can be used. With only 32,000 non-treatment households available and with the average fraction of about 9% of the non-treatment households that are likely to enroll (See Table 1), the prospective control group size can be estimated as 32,000*0.09 = 2,880, i.e., about a half of the treatment group size. Further, if we augment propensity score matching by matching with observable variables (i.e., electricity consumption before treatment in our case), the resultant control group would be even smaller. Our experimentation with such augmented nearest neighbor matching method resulted in a control group with only 245 households. This size discrepancy between a treatment and control groups lends one-by-one matching less reliable, since a household from a control group is matched to more than twenty treatment households on average[4]. Accordingly, we resorted to the interval sampling method [Caliendo & Kopeing 2008].

In the interval sampling method, the entire range of the propensity scores for class "enrolled" is divided into equal-probability intervals and the frequencies within each interval are calculated using the scores from the enrolled group. The control group is constructed from the non-treatment households using their calculated propensity score by matching their relative interval frequencies. The final PSM group included 2,586 households. In this method, we did not augment matching by propensity score with matching by the energy consumption either, because of the relatively small size of the candidate household pool.

We used our first score, i.e., the enrollment score, to construct the control group. Since we use the difference between the enrollment and not enrollment scores for classification, we also constructed a control group by matching the score differences. The resultant control group included 2,618 households of which 2,495 (96.5%) were also in the previous group. Due to this smallness of group difference and because the conventional PSM method operates with enrollment score, we decided to use the former group of 2,586 households as a PSM control group.

Energy Matched Control Group Construction

Due to popularity of non-experimental control group constructed by energy consumption matching [Allcott 2011], we decided to implement matching by energy consumption in the pre-treatment period to construct an alternative control group. We implemented a technique close to the M-nearest neighbor with replacement [Caliendo & Kopeing 2008] to construct the sample.

For each enrolled household, we selected M = 2 not-treatment households such that their squared Euclidian distance (integrated square difference in electricity consumption) is minimal over the pre-treatment year:

$$\sum_{t=1}^{365\cdot 24}\left(Y_{i,t}-Z_{j,t}\right)^2 \underset{j\in C}{\rightarrow} \min \tag{1}$$

In Equation (1), $Y_{i,t}$ is electricity consumption of enrolled household i during hour t and $Z_{j,t}$ is electricity consumption of not-treatment household j during the same hour. We applied a computational scheme based on Equation (1) to the 5,600 enrolled households and 32,000 non-treatment households. The resultant EM control group size was 2,300 households.

Observations from Non-Experimental Control Groups

In a separate paper [Zeifman 2014a], we studied possible dependence between the calculated household propensity score and average electricity consumption. We concluded that there were no dependence between the propensity score and household average electricity consumption. Accordingly,

---

[4] Note that we were given a small pool of candidate not-treatment households (32,000 as compared to 470,000 Program-eligible households). With an adequate candidate pool size, the nearest neighbor method can be implemented.

we did not expect the interval-based PSM control group, which was not augmented with matching by observables, to automatically match the treatment group by energy consumption. Indeed, while the average hourly electricity consumption of the treatment group (i.e., the enrolled households) over the pre-treatment year was 0.5957 kWh, that value for the PSM group was 0.7677 kWh, and for the EM group it was 0.5999 kWh.

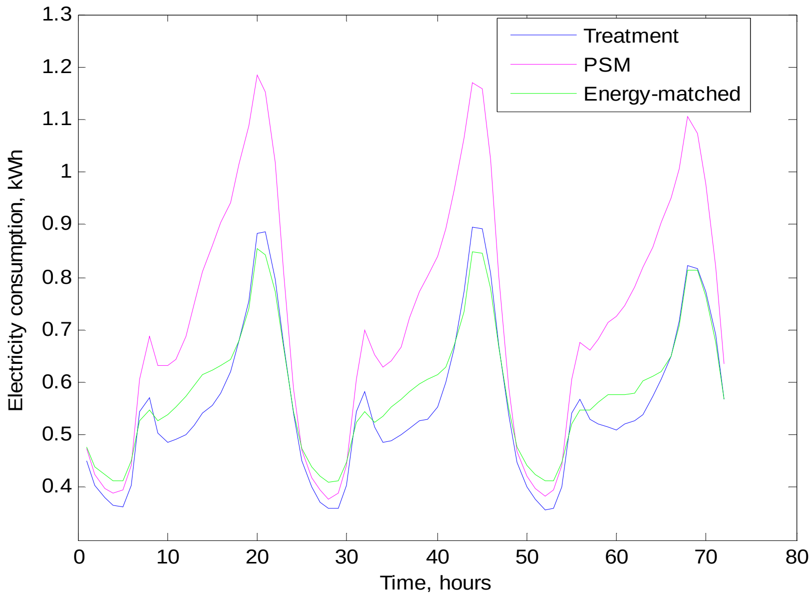

Figure 4. Hourly electricity consumption data for September 21-23, 2011, averaged over groups.

Since the average values differ considerably between the treatment group and the PSM control group, we can hypothesize that it is the details of electricity consumption that are related to the household propensity to enroll. Figure 4 partially supports this hypothesis. It can be observed in the Figure that the PSM group, though consuming more energy, somewhat better follows the patterns of the treatment group than the EM group. On the other hand, such a pattern does not always persist, so that it is difficult to argue what visible features are responsible for the household propensity to enroll. The nonlinearity of the selected NML algorithm implies that such features may not be directly observable.

Does the fact that the PSM control group consumes considerably more electricity than the treatment group invalidate it? First, had we been given a reasonably large pool of non-treatment households, (i.e., as large as the pool of the eligible households), we would have built a control group matched to the treatment group by both propensity to enroll and observable variables (see section PSM Control Group Construction). Second, it is possible to model different levels of energy consumption of the treatment and control groups in evaluation. We will explore this possibility in the next Section. Last, the main objective of this paper is to provide means to pinpoint households that are likely to enroll in behavior-based energy efficiency programs. For this objective, the difference in energy consumption is irrelevant.

Program Effect Estimation

In this Section, we give a brief example of the control group use for saving estimation along with limited results. To estimate the overall program saving effect and account for the discrepancy in electricity consumption of the treatment and PSM control group, we applied a panel data model with fixed effects [Braithwait, Hansen & Armstrong 2012] and used bootstrap-type simulations to estimate the saving confidence interval. To this end, in each simulation we select at random 400 households from a treatment group and 400 from a control group, and estimate the saving. We then repeat this process to obtain a statistical sample.

Figure 5 shows the distribution of saving percentage. For the PSM control group, the average saving is 4.94%, and the 90% confidence interval is [3.58% to 6.47%]. For the EM control group, the results are statistically different. The average saving is 1.69% with 90% confidence interval [-0.31% 4.10%], i.e., the saving is statistically insignificant at 5% significance level. These results also suggest that the distribution of saving percentage based on EM control group is approximately 50% wider than that based on the PSM group. This observation, together with the heavy tails of the EM-based distribution, are in line with Allcott's observations of very high variability of saving estimates calculated using the energy matched control groups [Allcott 2011].

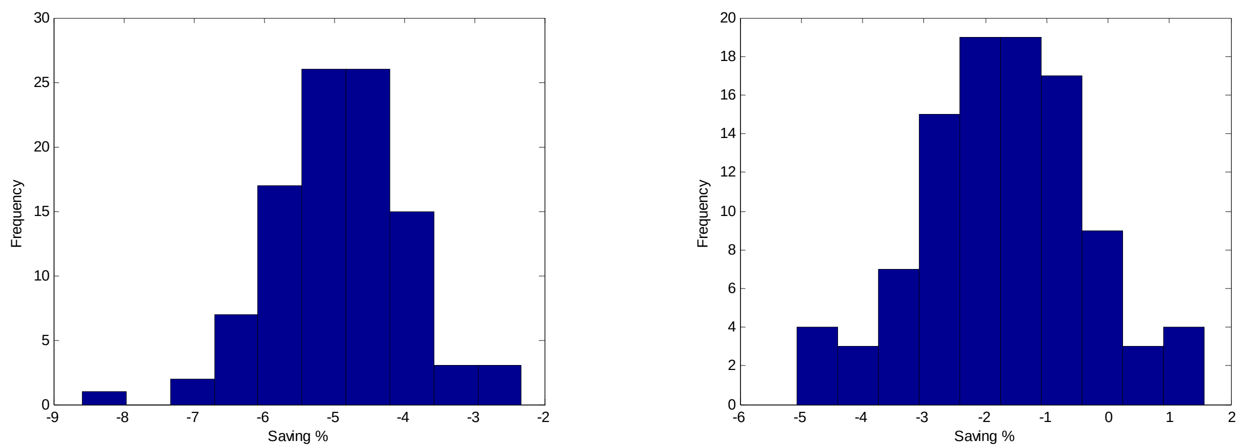

Figure 5. Distribution of saving percentage estimated by Equation (2) from 100 bootstrap simulations. Left: PSM control group used, Right: EM control group used.

Suggested Implementation

To use the proposed method for either non-experimental control group construction or identifying the households that are likely to enroll in an opt-in behavior based energy efficiency/demand response program[5], we suggest the following steps.

1. Obtain a sample of about a thousand households that enrolled in a current or past opt-in behavior based energy efficiency program. The sample needs to include interval electricity consumption data of each household collected over the same pre-treatment time window of at least one year.
2. Obtain a sample of about a thousand eligible for enrollment households that did not enroll in the same program, along with their interval data for the same period of time.
3. Train the algorithm using these two datasets. Although the algorithm seems to be complicated, its training can be performed by a "push button" in Excel worksheet and requires few seconds of computational time on a Windows machine. Estimate the goodness-of-fit by applying the algorithm to classification of households (likely or unlikely to enroll) and comparing the classification results with the actual enrollment data.
4. Obtain a pool of candidate households. These candidate households need to be as similar as possible to the households used for algorithm training in terms of geographic location and

[5] Whether the households that are likely to enroll in such programs are also likely to enroll in other energy efficiency programs, e.g., those that require home retrofits, is an open question that we need to address in future.

household type (e.g., detached homes or multifamily) and have the interval data available for the same period of time. The size of this pool is roughly ten times the desired prediction sample size or larger.

5. Using the trained algorithm and the interval data from the pool, calculate which candidate households are likely to enroll. In our case study, this prediction is another “push button” computer operation that requires few seconds of computational time for the pool size of 32,000.

Summary

We developed an easy-to-implement method for prediction of household propensity to enroll in an opt-in behavior based energy efficiency program using only information freely available to utilities. The method requires interval electricity consumption data collected over a year for a sample of 1,000-2,000 households that have enrolled in an opt-in behavior based energy efficiency program and a similar data set of households that have not enrolled. Once the main algorithm is fitted to these data, it can be used for enrollment prediction and/or quantitative characterization of propensity to enroll using the interval electricity data of other households. The prediction accuracy of our method, verified by using separate samples for algorithm fitting and for prediction, is shown to be about 90%.

It is known that matching by observable variables only in non-experimental control group construction does not eliminate the selection bias. Since such matching is not perfect, we deduce that the households with high propensity to enroll have specific energy spending patterns, which are lost in simple energy matching. Even though our method uses these patterns indirectly, they are still not directly observable even in the interval electricity consumption data.

One interesting finding of our case study is that the fraction of households with high calculated propensity to enroll can be as high as 9.6% whereas the actual rate of successfully recruited participants was only 1.2%. Since the predicted enrollment rate is in line with the average enrollment rate of 11% in several US opt-in behavior based energy efficiency programs [DOE 2013], we believe that our prediction of households that are likely to enroll is correct. Therefore, the actual enrollment rate could have been increased *sevenfold*, had the underlying Program used better advertisement. We have not attempted to experiment with better advertisement, however.

The proposed method is new, and there are numerous questions that need to be researched. For example, our method worked well for a given region/program. Will its performance be as good for a different region or program? Can the enrollment data from one program be used as a proxy for another program in the same region? Can our algorithm, trained on data from one region, be applied to data from another region? What are the requirements to the region (e.g., size, homogeneity)? Additional research work is needed to answer these questions.

Acknowledgement

The author is grateful to Peter Klint and Dr. Kurt Roth for insightful discussions and help with manuscript preparation and to Laura Schauer for review, valuable comments and suggestions. This work was supported by Ewing Marion Kauffman Foundation.

References

[Allcott 2011] H. Allcott, “Social norms and energy conservation.” Journal of Public Economics 95, pp. 1082-1095 (2011).

[Birt et al. 2012] B. J. Birt, et al. “Disaggregating categories of electrical energy end-use from whole-house hourly data.” Energy and Buildings 50, pp. 93-102 (2012).

[Braithwait, Hansen & Armstrong 2012] S. D. Braithwait, D. G. Hansen, & D. A. Armstrong, “2011 Impact Evaluation of San Diego Gas & Electric’s Peak-Time Rebate Pilot Program,” Ex-Post and Ex-Ante Report, Christensen Associates (2012).

[Caliendo & Kopeinig 2008] M. Caliendo, & S. Kopeinig, “Some Practical Guidance for the Implementation of Propensity Score Matching,” Journal of Economic Surveys 22, pp. 31–72 (2008).

[Dietterich 2000] T. G. Dietterich, "Ensemble methods in machine learning." In Multiple classifier systems, pp. 1-15. Springer Berlin Heidelberg (2000).

[DOE 2012] “Evaluation, Measurement, and Verification (EM&V) of Residential Behavior-Based Energy Efficiency Programs: Issues and Recommendations.” DOE, 2012.

[DOE 2013] “Analysis of Customer Enrollment Patterns in Time-Based Rate Programs – Initial Results from the SGIG Consumer Behavior Studies.” DOE, 2013.

[Edison 2012] “Utility-Scale Smart Meter Deployments, Plans, & Proposals.” Report by Institute for Electric Efficiency, the Edison foundation (2012).

[Harding & Hsiaw 2014] M. Harding & A. Hsiaw, “Goal setting and energy conservation.” Journal of Economic Behavior & Organization, in press (2014).

[Heckman et al. 1998] J. Heckman, H. Ichimura, J. Smith & P. Todd, “Characterizing Selection Bias Using Experimental Data”, Econometrica 66, pp. 1017-1098 (1998).

[Lechner 2000] M. Lechner, “An Evaluation of Public Sector Sponsored Continuous Vocational Training Programs in East Germany." Journal of Human Resources 35, pp. 347-375 (2000).

[Zeifman 2014a] M. Zeifman, “Smart meter data analytics: prediction of enrollment in residential energy efficiency programs.” Proceedings of IEEE Conference on Systems, Man and Cybernetics (SMC), pp. 3371-3374(2014).

[Zeifman 2014b] M. Zeifman, “Determining Probability of an Individual Consumer Enrolling in a Behavior-Based Energy Efficiency Program.” US patent application (2014).