Behavior and Energy Savings

Evidence from a Series of Experimental Interventions

Author

Matt Davis

Executive Summary

Energy efficiency is widely considered our greatest untapped energy resource. Swaths of research have identified a sizable "energy efficiency gap" consisting of un-pursued efficiency improvements that pay for themselves by reducing energy bills.[1] For various reasons, these investments are not being made, and as a result Americans waste billions of dollars in avoidable energy costs every year. Aggressively pursuing efficiency opportunities is the best and cheapest way to reduce greenhouse gas emissions, since they can reduce greenhouse gas emissions and earn a profit at the same time.

While there are many barriers to achieving this potential, at least part of the gap can be attributed to household behavior. Recent research has highlighted the many ways in which energy use is particularly prone to what traditional economics would deem "irrational" behavior. Electricity and heat are effectively invisible, their prices are delineated in abstract and unfamiliar units, and monthly billing ensures a temporal distance between usage and payment. Fortunately, the work of behavioral economists and scientists has provided some innovative strategies pursuing simple, low-cost savings by "nudging" people to make better energy-use decisions.

This report focuses on an innovative behavioral intervention that has been deployed in a wide variety of settings throughout America. Using data collected from 11 different utility service areas encompassing more than 750,000 households across the United States, this report estimates the effectiveness of "Home Energy Reports" (Reports) containing, among other things, information about how a household's electricity usage compares to that of its neighbors. In each case the Reports were delivered to a randomly selected subset of households, ensuring that any differences in electricity demand between the treatment and control groups can be attributed directly to the Reports' influence.

The results, summarized in the table below, convincingly demonstrate that this simple intervention is effective at reducing energy demand. Reports sent to a random subset of customers are shown to reduce energy demand by 1.8% on average, with the effectiveness of individual programs ranging from 0.9% to 2.9%. Furthermore, using a single readily available statistic – a household's average monthly energy usage before the intervention – can help identify the households most likely to show large reductions and increase average effectiveness to 6.5**%** overall. To get a sense of the magnitude of these effects, a simple extrapolation of the savings rate suggests that reducing residential electricity usage across the United States by 1.8% would save over 26,000 GWh of electricity, reduce greenhouse gas emissions by roughly 8.9 million metric tons of carbon dioxide per year – equal to the emissions from three 500-MW coal-fired power plants – and save households just over $3 billion dollars per year on their electric bills.[2]

[1] See Jaffe and Stavins (1994) and Dietz (2010) for more thorough discussions.
[2] According to the Energy Information Administration, residential electricity usage totaled 1,450,758 million kilowatt-hours in 2010, and the average national price is 11.58 ¢/kWh. A combined cycle natural gas power plant – which is a

Summary Of Results

Sample	Total Observations	Average Treatment Effect	“Targeted” Treatment Effect
1	2,355,000	-1.6%	-3.2%
2	3,108,000	-2.3%	-9.1%
3	1,159,000	-2.1%	-2.6%
4	496,000	-1.9%	-10.9%
5	730,000	-2.9%	-14.7%
6 (a)	1,487,000	-1.6%	-2.7%
6 (b)	1,923,000	-0.9%	-6.1%
7	3,011,000	-1.7%	-4.7%
8	2,644,000	-1.1%	-8.9%
9	504,000	-1.3%	-7.3%
10	3,351,000	-2.5%	-6.9%
11	1,460,000	-1.6%	-5.8%
Total / Wtd. Average:	*22,228,000*	*-1.8%*	-6.5%

Background

Engineers, environmental advocates, and a growing number of policy-makers agree that energy efficiency boasts a significant opportunity for energy savings. The most frequently cited estimate of the national potential speculates that achieving the full spectrum of Net Present Value (NPV)-positive efficiency improvements – that is, only the investments that pay for themselves – would reduce overall energy consumption by 23% and eliminate 1.1 gigatons of carbon dioxide emissions annually (Granade et. al. 2009). This points to a significant opportunity for energy savings that can be achieved without appealing to emissions fees or regulations.

The challenges to realizing this potential are many, but recent years have seen a surge in interest in behavioral “nudges” that can reduce energy demand at very low costs (see Allcott and

reasonable approximation of the “marginal” power plant in the United States – has a carbon intensity 0.34 metric tons per Megawatt-hour.

Mullainathan 2010 for a brief review). While a variety of programs have demonstrated promise, most suffer from a shortage of rigorous data demonstrating that observed or hypothesized reductions in energy use are both scalable and generalizable beyond a small sub-population. Even those that meet the "gold-standard" for program evaluation – randomized assignment of treatment – are subject the criticism of site specificity.

This issue is by no means unique to the world of energy-saving interventions. Indeed the problem at hand is the well-known issue of external validity: promising results in a single trial do not imply that we should expect identical results in another location with a different climate, culture and built environment, for instance. As such, it is vital to examine the drivers of differing program effectiveness (or to use the preferred academic terminology, "heterogeneous treatment effects"). Understanding these causes has important implications for both policy and program design as we consider how to best adjust and implement a portfolio of behavioral interventions.

However, there are many challenges to doing so. Evaluation of most behavioral interventions are limited to a single demonstration project, which inevitably lacks much of the richness and diversity necessary for extending results to the general population. One solution is repeated experiments in different settings, but repetition can be expensive, and the lack of "new" results can provide a disincentive for researchers to evaluate successive implementation of similar projects.

This study examines a large, geographically diverse data set to better understand how the results from a certain type of behavioral program can differ across geographical settings. While the sample encompasses a set of 11 utility service areas with significant diversity of income level, the analysis identifies a set of observable demographic variables that predictably influence the strength of the intervention's treatment effect. The analysis also suggests that using information about baseline energy usage to target reports at high-opportunity households can increase average effectiveness by more than three-fold. Similar targeting using a variety of other demographic indicators show smaller and unreliable gains, possibly due to issues of data availability.

Program Description

The data are gathered from a series of similar interventions in which various utilizes contracted with a private company – Opower – to provide a form of targeted energy-use feedback in the form of a Home Energy Report. Appendix 1 contains a sample report, but it is worth highlighting some of its salient features here:

Most notable (and original) is the peer energy comparison. Building on the results reported by Schultz et. al. (2007), the report displays a graphical comparison of a household's energy usage over the previous billing period to that of an "average neighbor," defined as the average of 100 nearby households, and an "efficient neighbor," representing the 20th percentile of the same sample. Those who use relatively less energy also receive injunctive feedback in the form of a smiley face, which has been demonstrated to reduce rebound effects in which low-usage consumers revert to the mean by increasing their energy usage (Schultz et. al. 2007).

While the peer comparison feedback is contained on every Report, the mailings also contain several other modules that vary from consumer to consumer and month to month. Examples include but are not limited to comparisons to the household's usage one year ago, a list of energy saving tips, and an estimate of how much money could be saved by reducing one's energy use to the average level. This analysis will not (indeed it cannot) disentangle the marginal effects of these specific elements; rather, the treatment effects discussed are the cumulative effects of the entire report.

Results from the Opower program have received considerable academic (and public) attention. Ayres et al (2009) analyze two separate interventions and report treatment effects of 1.2% and 2.1%; they also find that effectiveness is sensitive to the frequency with which households receive reports. Costa and Kahn(2010) examine a different kind of household heterogeneity – political views – and find that while "environmentalist" households respond by reducing their usage by 3.1%, "defiers"[3] may actually increase their usage by 0.7%; since the latter category account for a relatively small portion of their sample, however, the average household still shows a 2.1% reduction. Finally, Allcott (2010) analyzes a similar set of utilities to those described here; he reports treatment effects between 1.1% and 2.8% and demonstrates how statistical profiling can boost cost-effectiveness by 43%.

Data and Methods

The data are a very rich source of information about the intervention in question. For twelve interventions in eleven utility service areas,[4] EDF observed over 22 million electric meter-readings, dated between March 2008 and June 2010. Each of the programs in this sample benefits from experimental design; that is, each sample was randomly partitioned into a treatment and a control group before implementation. The data include both post-intervention electricity usage as well as at least twelve months of pre-intervention data for each household, allowing us to develop a robust baseline to measure program effectiveness and heterogeneity.

Data for each implementation are summarized in Table 1. The average treatment effect is calculated using a regression similar to that described in equation (1), but with the demographic interaction terms excluded (i.e. $X_{jit} = [1]$).

In addition to energy usage and indicators for whether a household has begun receiving Reports, we also observe a set of demographic variables. Summary statistics for the demographic variables included in our preferred specification are presented in Appendix 2.

[3] "Defiers" are defined as a Republican household that neither donates to environmental groups nor elects to pay a surcharge for renewably-sourced electricity, whereas "environmentalists" are registered Democrats who donate to advocacy groups and purchase green electricity.
[4] One utility has sent reports in two staggered implementations, each with its own treatment and control groups.

Table 1: Utility Summary Data

Code	Months of Treatment	Households	Observations	Average Treatment Effect
1	12	91,000	2,355,000	-1.6%
2	18	80,000	3,108,000	-2.3%
3	9	40,000	1,159,000	-2.1%
4	9	35,000	496,000	-1.9%
5	15	18,000	730,263	-2.9%
6 (a)	10	50,000	1,487,000	-1.6%
6 (b)	6	108,000	1,923,000	-0.9%
7	22	84,000	3,011,000	-1.7%
8	18	70,000	2,644,000	-1.1%
9	10	40,000	504,000	-1.3%
10	28	85,000	3,351,000	-2.5%
11	8	70,000	1,460,000	-1.6%
	Total/Average:	*771,000*	*22,228,000*	*-1.8%*

The identification procedure is very simple, given the randomized assignment of treatment throughout the sample. Since assignment into the treatment group is uncorrelated with other regressors, a simple differences-in-differences regression provides unbiased estimation of both the average treatment effect and the associated interaction terms.

The regression model is expressed mathematically as follows:

$$Usage_{it} = \beta_0 + \beta_1 Post_{it} + \beta_2 Post_{it} * Treatment_i\left[\sum_j \alpha_j * X_{jit}\right] + \sum_k \gamma_k Z_{kit} + f_i + \epsilon_{it} \quad (1)$$

"Post" is an indicator variable set equal to one if household i has started receiving Home Energy Reports prior to date t, while "Treatment" indicates whether household i is in the treatment group.[5] Hence, the coefficient on the interaction of these terms estimates the average treatment effect. Usage is normalized in all regressions, so we interpret this effect as a percentage.

X_{jit} is a column vector of household demographics that we then interact with the treatment effect; the corresponding regression coefficients– denoted by α_j – estimate the marginal effects of different demographic variables upon the magnitude of the treatment effect and are the main coefficients of interest. Z is a vector consisting of month dummy variables and heating and cooling degree days to control for variations in weather; while the experimental design ensures consistent estimation of the treatment effect regardless of whether we include these controls, they are significant predictors of household energy use and therefore can increase the power of our estimators. The error term is partitioned into a household-level fixed effect f and a random error term ϵ that by construction has mean zero and is uncorrelated with all regressors.

Estimation of this model is straightforward using traditional regression techniques. Standard errors are clustered at the household level to ensure that estimates are robust to autocorrelation.

Results

Appendix 3 displays the regression results of equation (1) for each of the twelve utility programs. For this first set of regressions, X_{jit} – the vector of variables whose interactions with the treatment effect we hope to study – includes a full set of household demographic variables, allowing us to examine how treatment effects vary with baseline energy usage, house square footage, the number of household occupants, the head of household's age, the age of the physical house, household income, and two dummy variables indicating whether the household is a single-family home and whether it is rented.

One immediately apparent result is the clear presence of unobserved confounding variables. While observational data can predict significant heterogeneity *within* each utility, the same variables do not have identical effects – or even the same sign – across different utilities. This suggests the presence of unobserved variables (such as political affiliation, baseline home efficiency, etc.) that bear influence on the Reports' effectiveness and are correlated with the observed variables. In plain language, while it is clear that effectiveness varies across different types of households, the observable data do not explain enough of this variation to reliably explain why certain samples show higher or lower average treatment effects.

From an econometric point of view, this result forces limits the analysis to utility-level regressions. While it is tempting to combine all of the results and estimate a cumulative model for the entire sample, such a process is almost certainly subject to biased parameter estimates. As such, general results are limited to qualitative conclusions based on the collection of utility-level estimates, with the caveat that what is predictive in a specific setting is not causal in a general setting.

To simplify these results, Appendix 4 presents the results of a shorter regression that uses only the four demographic variables listed in Table 2. These variables were selected because they display the most consistent relationship with treatment effectiveness; they are statistically significant at the 90% confidence level in at least 2/3 of the regressions in which they appear,

[5] Households have the opportunity to opt out of treatment, though in practice a very small fraction (<1.5%) do so. As such, treatment effects should be interpreted as "intent-to-treat" rather than "treatment-on-treated."

and a sizeable majority of those estimates have the same sign. The other four variables omitted on this specification – house vintage, income, rental status, and single-family status – are either rarely statistically significant or do not display any clear directional pattern across utilities.

While the full details of the regressions are left to the Appendix, Table 2 provides a high-level summary of how these observable variables interact with the treatment effect. The values in columns 2 and 3 denote the number of utilities for which an increase in the given variable is associated with either an increase or a decrease in the magnitude of the treatment effect. This quick gloss suggests that, for the majority of utilities in the sample, the Reports' effectiveness is greater in households with:

- Higher baseline usage
- Smaller square footage
- Fewer Occupants
- Older heads of household

Table 2: Variables' Influence on Treatment Effects

	Increases Effectiveness	Decreases Effectiveness	Median Value	Effect of Standard Deviation Increase
Baseline Usage (kWh/day)	9	2	-.00482	-6.81%
House Size (1,000 ft^2)	2	9	0.00930	+0.76%
Number of Occupants	1	6	.00615	+0.82%
Head of Household's Age (years)	11	0	-.000704	-1.54%

The median values reported here can be interpreted as percentages. For instance, as baseline energy increases by one kWh/day, we would expect treatment effects in that household to be roughly half of a percentage point higher. Similarly, effectiveness is expected to decrease by roughly 0.6 percentage points with each additional household member and 0.9 percentage points with each additional 1,000 square feet of housing, while increasing by 0.7 percentage points as the head of household's age increases by ten years.

To express these magnitudes in similar units, the final column details the expected change in treatment effectiveness after a one-standard-deviation increase in the variable in question. As previous studies have suggested, baseline usage proves to be the most significant driver of heterogeneity. We predict that a household with a one-standard-deviation increase in baseline usage would show an increase in effectiveness of 6.8% points – more than three times the mean effect in the sample.

The results described thus far point to several factors that can be used to predict treatment effectiveness for different households. Taken at face value, this information can be used to identify households that are likely to show above-average responses to the Reports, thereby increasing average effects and cost-effectiveness.

Table 3 considers two such "targeting" strategies. Both use the predictions generated by a regression of equation 1 to identify the expected treatment effects for every household in the sample. This allows one to select the half of the sample with the greatest anticipated effects.[6] Since baseline usage explains the most variance of the variables considered and is also likely to be the most easily observable for a utility, the middle column displays estimated treatment effects for a regression model that uses only baseline usage as an interaction term. The results in the rightmost column use a model that includes the three additional variables as well. Untargeted effects are the average treatment effects reported in Table 1 and are included for ease of comparison.

Table 3: Treatment Effects With Targeting

	No-Targeting	Targeting Using Baseline Usage	Targeting Using All Variables
1	-1.6%	-3.2%	-4.3%
2	-2.3%	-9.1%	-8.2%
3	-2.1%	-2.6%	-6.4%
4	-1.9%	-10.9%	-12.6%
5	-2.9%	-14.7%	-12.8%
6 (a)	-1.6%	-2.7%	-2.7%
6 (b)	-0.9%	-6.1%	-5.9%
7	-1.7%	-4.7%	-4.9%
8	-1.1%	-8.9%	-9.9%
9	-1.3%	-7.3%	-7.7%
10	-2.5%	-6.9%	-7.5%
11	-1.6%	-5.8%	-5.8%

The gains to profiling using baseline usage alone are substantial. Using this simple metric to identify high-potential households boosts the average treatment effects from a cross-program average of 1.8% to 6.5%.

Integrating demographic data, however, shows less promise. In seven cases observed data can improve the treatment effect relative to using baseline usage alone, but in many cases the additional gain is small, or even negative. The negative examples are likely symptoms of missing data issues, as households for which we do not observe, for example, square footage, are

[6] Of course, a real program would not have access to its own performance data before implementation. However, we can imagine a program designer sending Reports to a small, randomly selected subset of households, analyzing the results, and using those conclusions to target implementation.

dropped from the final regressions. If data are scarcer among high potential users, we might expect the yield from conditioning to decrease.

For program planners, the results point to the power of targeting treatment at high-usage households. This data should be readily available to utilities, and it does an excellent job identifying which households have the greatest potential reductions.

Conclusions

The results of this paper demonstrate that a simple behavioral intervention demonstrates consistent savings across a wide range of utilities. The programs in the sample serve urban and rural areas, the geography spans east coast, west coast, and Midwestern states, and the service providers include from investor-owned utilities, coops, and municipal utilities.

Likewise, there are some important differences in how households respond to Home Energy Reports. While the explanatory power is not sufficient to predict out-of-sample performance, our analysis shows that conditioning on one readily observable piece of data could increase average treatment effects by more than three-fold on average.

Future research should focus on performing similar analyses of other approaches to reducing energy demand. Where heterogeneity can be effectively and rigorously studied, there is considerable value in understanding how different households respond to different treatments. This knowledge is essential in developing and deploying the optimal set of efficiency programs. Targeting the right programs at the right households ensures that energy-saving resources are put to their best use.

References

Allcott, Hunt. 2011. “Social Norms and Energy Conservation.” *Journal of Public Economics*, In Press, Corrected Proof, Available online 21 March 2011, ISSN 0047-2727.

Allcott, Hunt and Sendhil Mullainathan. 2010. “Behavior and Energy Policy.” *Science*. 327 (5970), pp. 1204-1205.

Ayres, Ian, Sophie Raseman, and Alice Shih. 2009. “Evidence from Two Large Field Experiments that Peer Comparison Feedback can Reduce Residential Energy Usage.” 5th Annual Conference on Empirical Legal Studies Paper. Available at SSRN: http://ssrn.com/abstract=1434950

Costa, Dora and Matthew Kahn. 2010 “Energy Conservation ‘Nudges’ and Environmentalist Ideology: Evidence from a Randomized Residential Electricity Field Experiment.” NBER Working Paper 15939.

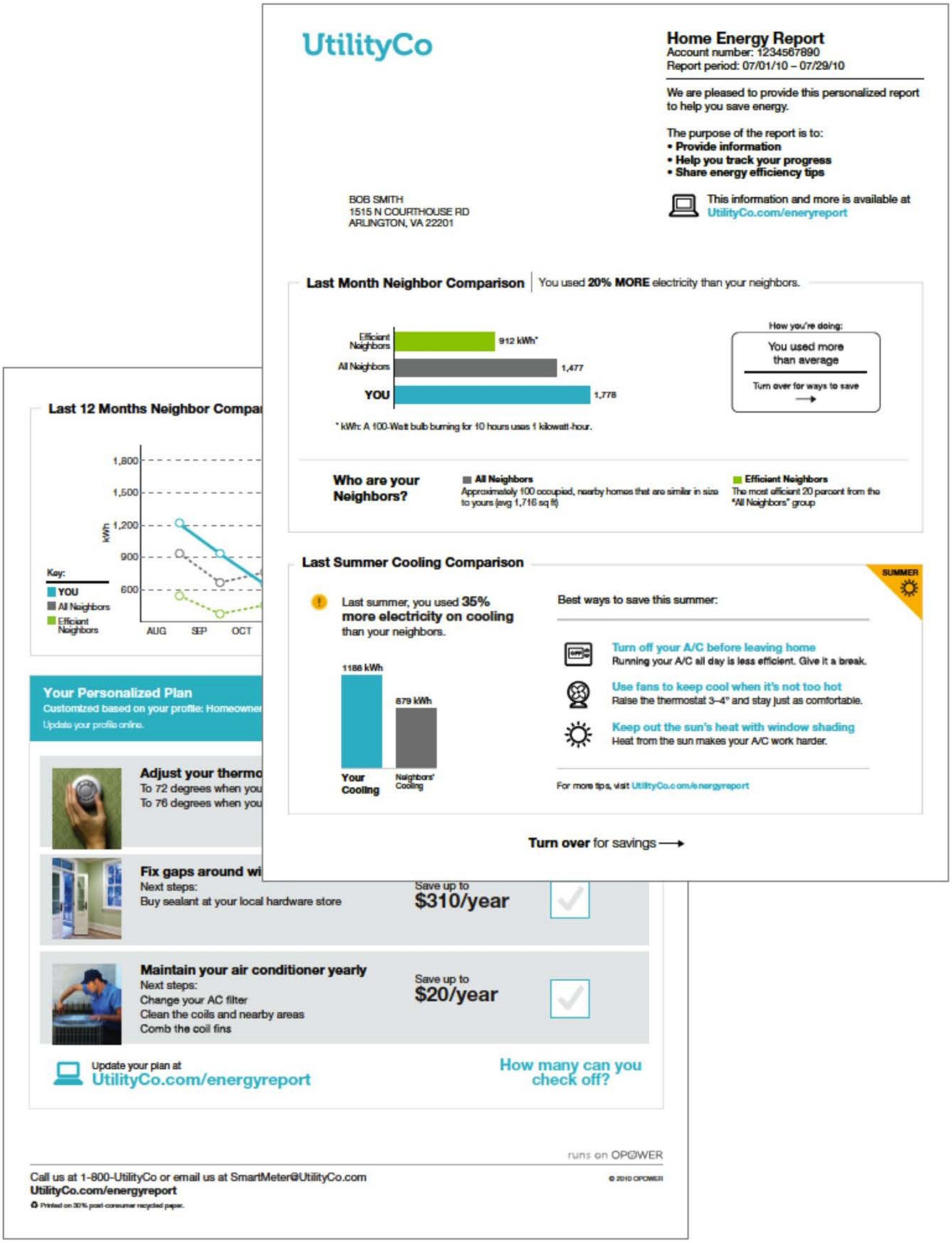
Dietz, Thomas. 2010. “Narrowing the U.S. Energy Efficiency Gap.” *Proceedings of the National Academy of Sciences***,** 107 (37) pp. 16007-16008.

Granade, Hannah Choi, Jon Creyts, Anton Derkach, Philip Farese, Scott Nyquist, Ken Ostrowski. 2009. “Unlocking Energy Efficiency in the U.S. Economy.” McKinsey & Company report, available at http://www.mckinsey.com/en/Client_Service/Electric_Power_and_Natural_Gas/Latest_thinking/~/media/McKinsey/dotcom/client_service/EPNG/PDFs/Unlocking%20energy%20efficiency/US_energy_efficiency_full_report.ashx

Jaffe, Adam B. and Robert N. Stavins. 1994. “The energy-efficiency gap: What does it mean?” *Energy Policy* 22 (10), pp. 804-810.

Schultz, Wesley, Jessica Nolan, Robert Cialdini, Noah Goldstein, and Vladas Griskevicius. 2007. “The Constructive, Destructive, and Reconstructive Power of Social Norms.” *Psychological Science* 18 (5), pp. 429-434

Appendix 1: Sample Home Energy Report



Appendix 2: Summary Statistics

Utility Code	House Size (ft^2)	Age: Head of Household	Number of Occupants	Baseline Usage (kwh/day)
1	2,397.67	54.48	4.69	42.51
	(1,385.87)	(12.38)	(1.70)	(25.75)
2	1,696.39	49.88	2.62	30.66
	(536.50)	(13.20)	(1.32)	(16.99)
3	2,009.82	50.54	4.26	25.42
	(779.32)	(14.77)	(0.63)	(13.37)
4	1,686.03	55.82	4.14	18.46
	(718.48)	(12.89)	(0.47)	(10.82)
5	1,282.23	59.47	3.11	39.93
	(535.04)	(12.95)	(1.53)	(29.05)
6 (a)	2,033.59	55.15	4.46	30.17
	(846.25)	(12.22)	(0.79)	(14.78)
6 (b)	1,784.91	53.62	4.49	33.82
	(702.73)	(11.94)	(0.82)	(13.84)
7	2,167.73	55.65	2.21	29.87
	(652.00)	(13.92)	(1.07)	(13.39)
8	1,869.57	61.57	(unobserved)	29.55
	(769.74)	(15.76)		(21.24)
9	1,831.96	56.22	1.88	36.30
	(769.57)	(12.59)	(0.99)	(16.89)
10	1,746.27	54.63	1.93	30.66
	(596.65)	(13.26)	(1.00)	(15.21)
11	1,634.14	55.64	4.35	27.98
	(645.08)	(12.42)	(0.77)	(11.47)

Appendix 3: Regression Results (Full Model)

	(1)	(2)	(3)	(4)	(5)	(6a)
Past	-0.0116***	-0.00978 ***	0.0301***	0.00824*	-0.0791***	-0.00138
	(0.00156)	(0.00114)	(0.00470)	(0.00483)	(0.00827)	(0.00135)
Treatment*Past	0.143***	0.155***	0.104***	0.186***	0.149***	0.0179
	(0.0245)	(0.00734)	(0.0291)	(0.0343)	(0.0497)	(0.0140)
T*P*Baseline	-0.00243 ***	-0.00613 ***	0.00132	-0.0133***	-0.00625 ***	0.000756***
	(0.000257)	(0.000159)	(0.000807)	(0.000446)	(0.000486)	(0.000175)
T*P*sqft	8.82e-06***	1.78e-05 ***	-3.47e-05***	3.96e-05***	6.84e-05***	5.12e-06**
	(2.12e-06)	(2.63e-06)	(6.45e-06)	(5.92e-06)	(2.29e-05)	(2.11e-06)
T*P*num_occ	0.00879***	0.00351***	-0.00730	0.00548	0.0126***	0.00225
	(0.00185)	(0.00100)	(0.00485)	(0.00593)	(0.00363)	(0.00170)
T*P*occ_age	-0.000339 ***	-0.00120 ***	-0.000651 ***	-0.000594**	-0.000724	-0.000738 ***
	(0.000130)	(8.75e-05)	(0.000213)	(0.000248)	(0.000466)	(0.000109)
T*P*(Rent)	-0.0318*	0.0241***	0.0204*	-0.00570	0.0225	-0.00496
	(0.0185)	(0.00574)	(0.0109)	(0.0134)	(0.0224)	(0.00958)
T*P*(1-family)	-0.0959***	0.0230***	-0.0309**	0.0253***		
	(0.0193)	(0.00497)	(0.0133)	(0.00862)		
T*P*income	-0.000559	0.000629	-4.05e-07 ***	-7.94e-08	-1.38e-07	-0.00422***
	(0.00147)	(0.000390)	(8.59e-08)	(9.10e-08)	(1.85e-07)	(0.00112)
T*P*house_age	8.46e-05	-0.000142 ***	0.00136***	0.000549 ***	-5.60e-05	-2.24e-05
	(7.46e-05)	(4.26e-05)	(0.000160)	(0.000145)	(8.23e-05)	(3.01e-05)
Observations	966,428	1,697,039	550,164	152,994	195,925	1,056,865
R-squared	0.208	0.226	0.061	0.304	0.119	0.177
Number of id	37,417	42,651	18,466	10,370	4,805	32,979

*Standard errors clustered by household reported in parentheses. In addition to variables reported here, regressions include controls for heating degree days, cooling degree days, and month dummy variables. *, **, and *** denote significance at 90%, 95%, and 99% confidence levels, respectively.*

Appendix 3, Continued

	(7)	(8)	(9)	(10)	(11)	(12)
Past	0.0291***	-0.0187***	-0.0323***	-0.00741***	-0.0212***	-0.0183***
	(0.00120)	(0.00101)	(0.00247)	(0.00274)	(0.000963)	(0.00126)
Treatment*Past	0.163***	0.0535***	0.152***	-0.150***	0.0832***	0.170***
	(0.00807)	(0.00852)	(0.0158)	(0.0213)	(0.00726)	(0.0106)
T*P*Baseline	-0.00577***	-0.00316***	-0.00590***	0.00532***	-0.00460 ***	-0.00502 ***
	(0.000109)	(0.000160)	(0.000271)	(0.000424)	(0.000130)	(0.000186)
T*P*sqft	1.02e-05***	1.17e-05***	3.58e-05***	-2.73e-05***	4.17e-05***	2.17e-06
	(1.29e-06)	(2.18e-06)	(5.59e-06)	(4.42e-06)	(2.55e-06)	(2.31e-06)
T*P*num_occ	0.00942***	0.00188*			0.00826 ***	0.00462***
	(0.000850)	(0.00108)			(0.00124)	(0.00150)
T*P*occ_age	-0.000940 ***	-0.000638 ***	-0.000587 ***	9.98e-06	-0.00115 ***	-0.000852 ***
	(6.16e-05)	(9.00e-05)	(0.000146)	(0.000224)	(8.87e-05)	(9.39e-05)
T*P*(Rent)	0.00176	-0.00670	0.0267*	-0.00500	-0.0211	0.0178
	(0.00396)	(0.0146)	(0.0157)	(0.0154)	(0.0141)	(0.0154)
T*P*(1-family)	-0.00621***		-0.00581			
	(0.00197)		(0.00423)			
T*P*income	1.52e-07***	1.53e-08	-1.26e-08	1.62e-08	-4.12e-09	-5.49e-08*
	(2.14e-08)	(3.16e-08)	(4.11e-08)	(6.50e-08)	(2.99e-08)	(3.20e-08)
T*P*house_age	-4.57e-05***	0.000848 ***	0.000396**	-0.000166	7.35e-05*	-0.000352 ***
	(1.27e-05)	(8.65e-05)	(0.000179)	(0.000104)	(3.81e-05)	(4.06e-05)
Observations	1,233,186	1,803,374	927,927	289,838	1,968,206	843,826
R-squared	0.188	0.260	0.590	0.333	0.261	0.171
Number of id	68,717	50,112	22,776	22,535	47,004	40,194

*Standard errors clustered by household reported in parentheses. In addition to variables reported here, regressions include controls for heating degree days, cooling degree days, and month dummy variables. *, **, and *** denote significance at 90%, 95%, and 99% confidence levels, respectively.*

Appendix 4: Regression Results (Restricted Model)

	(1)	(2)	(3)	(4)	(5)	(6b)
Past	-0.0118***	-0.00978 ***	0.0304***	0.00826*	- 0.0854***	0.0293***
	(0.00156)	(0.00114)	(0.00470)	(0.00483)	(0.00891)	(0.00120)
Treatment * Past	0.0396***	0.174***	0.159***	0.218***	0.197***	0.162***
	(0.0150)	(0.00662)	(0.0256)	(0.0335)	(0.0390)	(0.00643)
T*P*Baseline	-0.00221***	-0.00610 ***	0.00103	-0.0130 ***	-0.00600 ***	-0.00574 ***
	(0.000264)	(0.000156)	(0.000787)	(0.000432)	(0.00040 4)	(0.000109)
T*P*sqft	6.49e-06***	2.24e-05***	-5.67e-05 ***	3.35e-05 ***	2.19e-05*	1.21e-05***
	(1.87e-06)	(2.31e-06)	(6.55e-06)	(5.86e-06)	(1.18e-05)	(1.38e-06)
T*P* Num_Occ	0.00873***	0.00387** *	-0.0149***	0.00757	0.0117***	0.0107***
	(0.00184)	(0.000977)	(0.00487)	(0.00582)	(0.00320)	(0.00084 3)
T*P*Age_Occ	-0.000246*	- 0.00131***	-0.000668 ***	-0.000447 *	- 0.000802 *	- 0.000946 ***
	(0.000127)	(8.06e-05)	(0.000212)	(0.000243)	(0.000416)	(6.12e-05)
Observations	1,003,863	1,697,039	551,873	153,186	245,319	1,234,170
R-squared	0.207	0.226	0.059	0.304	0.103	0.188
Number of id	38,859	42,651	18,523	10,383	6,018	68,772

*Standard errors clustered by household reported in parentheses. In addition to variables reported here, regressions include controls for heating degree days, cooling degree days, and month dummy variables. *, **, and *** denote significance at 90%, 95%, and 99% confidence levels, respectively.*

Appendix 4, Continued

Utility Code:	(6a)	(7)	(8)	(9)	(10)	(11)
Past	-0.00139	-0.0187***	-0.00740 ***	-0.0320***	-0.0198***	-0.0183***
	(0.00135)	(0.00101)	(0.00274)	(0.00235)	(0.000942)	(0.00126)
Treatment * Past	-0.00883	0.0829***	-0.162***	0.162***	0.0914***	0.146***
	(0.0114)	(0.00750)	(0.0192)	(0.0127)	(0.00651)	(0.0103)
T*P*Baseline	0.000735 ***	-0.00304 ***	0.00534 ***	-0.00582 ***	-0.00456 ***	-0.00508 ***
	(0.000174)	(0.000160)	(0.000424)	(0.000232)	(0.000125)	(0.000185)
T*P*sqft	3.35e-06	4.69e-06**	-2.62e-05 ***	2.96e-05***	4.12e-05 ***	5.23e-06**
	(2.04e-06)	(1.91e-06)	(4.22e-06)	(4.60e-06)	(2.24e-06)	(2.03e-06)
T*P*Num_Occ	0.00127	0.00163			0.00758***	0.00473***
	(0.00169)	(0.00107)			(0.00121)	(0.00147)
T*P*Age_Occ	-0.000740 ***	-0.000424 ***	1.35e-06	-0.000525 ***	-0.00122 ***	-0.000893 ***
	(0.000107)	(8.01e-05)	(0.000211)	(0.000133)	(8.44e-05)	(9.36e-05)
Observations	1,058,413	1,803,374	289,838	1,060,356	2,108,326	843,868
R-squared	0.177	0.260	0.333	0.574	0.259	0.171
Number of id	33,027	50,112	22,535	26,033	50,359	40,196

*Standard errors clustered by household reported in parentheses. In addition to variables reported here, regressions include controls for heating degree days, cooling degree days, and month dummy variables. *, **, and *** denote significance at 90%, 95%, and 99% confidence levels, respectively.*

This report was written by Matt Davis, Research Fellow at Environmental Defense Fund.

Environmental Defense Fund, 257 Park Avenue South, New York, NY, 10010