Heterogeneity Analysis Report, EPC-14-026

Judson Boomhower Lucas Davis*

November 2016

Abstract

This Heterogeneity Analysis Report describes results from the third phase of "Evaluating Take-Up and Savings in a Large-Scale Air Conditioner Replacement Program", EPC-14-026. The overall objective of the project is to understand take-up and savings in Southern California Edison's *Quality Installation Program*, a large-scale central air conditioner replacement program. This report describes estimates from the third phase of the project aimed at describing how take-up and savings vary across customers. Using the same empirical difference-in-differences framework used in the second phase, we expand the analysis in this phase to examine heterogeneity. The results show large variation in energy savings between mild, warm, and hot areas of Southern California Edison territory. On average, program participants in hot areas (Climate Zones 13, 14, and 15) save 1,100+ kilowatt hours annually, compared to 300 kilowatt hours annually in warm areas (Climate Zones 9 and 10), and approximately zero average savings in mild areas (Climate Zones 6, 8, and 16). We also test for variation in savings between locations with different levels of household income, education, racial makeup, and household size, but these factors prove to be much less significant than climate. These demographic factors strongly influence take-up, however. For example, we find that higher-income households are approximately twice as likely to participate in the program as lower-income households. These results have important implications for the cost-effectiveness of the program as well as for the potential for better targeting of the program to increase benefits. Lower-income households in hot areas would appear to be a particularly valuable potential target.

*(Boomhower) Stanford Institute for Economic Policy Research, Stanford University. Email: boomhower@stanford.edu. (Davis) Haas School of Business, University of California, Berkeley. Email: ldavis@haas.berkeley.edu. We are grateful to Andrew Campbell and seminar participants at Stanford University, the AERE annual meeting, UC Berkeley, and Lawrence Berkeley National Lab for helpful comments and to Ellen Lin and Matt Woerman for excellent research assistance. This research was supported by the California Energy Commission under EPC-14-026.

1 Introduction

This report describes estimates from the third phase of the project EPC-14-026 aimed at describing how take-up and energy savings vary across customers. We first briefly describe our empirical strategy which follows closely from Boomhower and Davis (2016). Using the same empirical difference-in-differences framework used in the second phase of EPC-14-026, we expand the analysis in this phase to examine heterogeneity. We do this first with a series of "paired" regressions, and then with a single regression which pits the different types of heterogeneity against each other in attempt to disentangle which factor is most important.

We then describe results. Most significantly, the results show large variation in energy savings between mild, warm, and hot areas of Southern California Edison territory. On average, program participants in hot areas (Climate Zones 13, 14, and 15) save 1,100+ kilowatt hours annually, compared to 300 kilowatt hours annually in warm areas (Climate Zones 9 and 10), and approximately zero average savings in mild areas (Climate Zones 6, 8, and 16). This central finding comes through in both the "paired" and single regression approaches.

We also test for variation in savings between locations with different levels of household income, education, racial makeup, and household size. These factors prove to be much less significant than climate, however. In particular, once we include climate, none of these factors prove to have a large effect on energy savings. The coefficient estimates for these non-climate factors also don't have any consistent pattern, further underscoring the main finding that climate seems to be the most important form of heterogeneity.

We conclude by discussing the potential implications of these results for the cost-effectiveness of the program as well as for the potential for better targeting of the program to increase benefits. Even though energy savings vary relatively little with demographics, these demographic factors do strongly influence take-up. For example, we find that higher-income households are approximately twice as likely to participate in the program as lower-income households. Underserved groups in hot climate areas would thus appear to be a particularly valuable potential target for future programs.

2 Empirical Strategy

Our estimating equation is essentially identical to the equation used in Boomhower and Davis (2016). The response variable is electricity consumption measured in kilowatt hours. The explanatory variable of interest is 1[New Air Conditioner] an indicator variable equal to one for participating households after they have replaced their air conditioner through the *Quality Installation Program*. All regressions include our complete set of fixed effects including household by hour-of-day by month-of-year, and week-of-sample by hour-of-day fixed effects, and thus identify savings via difference-in-differences.

The innovation in this third phase of EPC-14-026 is to look not only at average electricity savings, but also at how savings vary across participants. We do this first with a series of paired regressions in which we simply divide our sample into different subsets. For example, participants for which annual household income in that zip code is above or below median annual household income.

In addition to these paired regressions we also estimate a single regression which pits the different types of heterogeneity against each other in attempt to disentangle which factor is most important. In particular we estimate a model with 1[New Air Conditioner] interacted with household income, educational attainment, climate, and the other factors.

We do not have household-level demographic information. Instead, we use Census data from a private vendor called *Geolytics* that has household income, educational attainment, race, and household size available at the Zip9 level. We imputed, therefore, for each household demographic variables equal to the average characteristics for their Zip9.

3 Results

3.1 Paired Regressions

Table 1 reports estimates from the paired regressions. In addition to reporting estimates for energy savings, this table reports the the take-up rate for each group. These take-up rates are of significant independent interest because they can indicate how successful the program has been on reaching different categories of households.

The structure of Table 1 is as follows. Column (1) reports the number of households in each category who participated in the program. Column (2) reports the participation rate for that group. Finally, column (3) reports the average annual electricity savings (in kWh) for program participants in that group.

We start in Panel A by reporting estimates for all participating households. Overall, there were 7,284 households in our data who participated in the *Quality Installation Program*. This is about one-tenth of one percent of all 4.9 million residential customers in PG&E territory. And, on average, program participants saved 398.6 kilowatt hours annually after replacing their central air conditioners. These results are identical to the baseline results from the second phase of EPC-14-026.

Panel B breaks participants into two categories on the basis of annual household income. Take-up is about twice as high for higher-income households. The pattern of high uptake among higher-income households has been previously noted with other similar programs (see, e.g. Borenstein and Davis, 2015) and is of great policy relevance. Average energy savings are larger in the lower-income category. Having now done additional analysis, however, we believe that this pattern is almost entirely due to climate. Lower-income households come disproportionately from the hot climate zones. With these paired comparisons climate is not controlled for, so these differences in Table 1 can reflect climate and other "lurking variables" rather than income itself. Indeed, when we run a single regression below, it turns out that climate is a much more important factor and income.

Panel C looks at educational attainment. Both take-up and savings are positively correlated with educational attainment. It is perhaps surprising that the income and educational attainment results appear to go in opposite directions. We've examined a map of the different categories (Figure 1) and although income and educational attainment are strongly correlated, they are not perfectly correlated. In particular there are a large number of "low" income, high education zip codes in the Palm Spring area, potentially indicating retirees.

Panel D examines climate zones. Take-up is near zero in the mild zones, and at .23% in the warm and hot areas. Savings are strongly correlated with climate zone. In the mild zones, the point estimate for savings is near zero and not statistically significant. Savings are larger in warm areas, and then much larger again in the hot areas, with average savings of almost 1,200 kilowatt hours annually. We think these are some of the most interesting results of the heterogeneity analysis, and of direct policy relevance.

Panels E and F look by race and household size. Overall we find smaller savings in zip codes which are more than 50% non-white and with larger households. Also notable is that the take-up rate is considerably smaller in zip codes that are more than 50% non-white.

As the truism goes, however, correlation is not causation. All of these patterns could reflect correlation between these characteristics and lurking variables. For example, the estimates for household income could be reflecting correlation between income and climate zones, rather than any true causal impact of income. The large variation in savings by climate zone, in particular, we believe is driving most of these results as becomes clearer in the next subsection.

3.2 Single Regression

Table 2 examines the same heterogeneity but with a single regression which pits the different types of heterogeneity against each other in attempt to disentangle which factor is most important. In particular we report coefficient estimates and standard errors from a single least squares regression. As with Table 1, the dependent variable is average hourly electricity consumption and the regression includes household by hour-of-day by month-of-year, and week-of-sample by climate zone fixed effects.

The difference is that the specification includes a set of interaction between 1[New Air Conditioner] and six different heteogeneous factors listed in the row headings. Program participants belong to different categories for these different factors (e.g. above median income, warm climate zone, more than three people per household), and the single regression is an attempt to tease out which factors are most significant.

The results are very interesting. Most significantly, the results show large variation in energy savings between mild, warm, and hot areas of Southern California Edison territory. On average, program participants in hot areas (Climate Zones 13, 14, and 15) save 1,166 kilowatt hours annually, compared to 378 kilowatt hours annually in warm areas (Climate Zones 9 and 10), and approximately zero average savings in mild areas (Climate Zones 6, 8, and 16). This central finding comes through in the single regression approaches, even after allowing for heterogeneity in these other factors.

The demographic factors are much less significant. While there were large differences for high- and low-income in the paired comparison, here the coefficient on high income is relatively small (65). Thus, the single regression provides strong evidence that the difference observed in the paired comparison is driven by climate, rather than by the "true" effect of income. Results are similarly small, and of inconsistent pattern for the other demographic factors.

4 Concluding Comments

The main finding from the third phase of "Evaluating Take-Up and Savings in a Large-Scale Air Conditioner Replacement Program", EPC-14-026, is that climate is the most important form of heterogeneity. On average, program participants in hot areas (Climate Zones 13, 14, and 15) save 1,100+ kilowatt hours annually, compared to 300 kilowatt hours annually in warm areas (Climate Zones 9 and 10), and approximately zero average savings in mild areas (Climate Zones 6, 8, and 16).

These results imply that the *Quality Installation Program* is likely to be most cost-effective in the hot areas of Southern California Edison's territory. At some level, this makes intuitive sense. Where air conditioners are mostly heavily used, energy-efficiency gains are most valuable and translate into the largest total savings in kilowatt hours.

Demographic factors strongly influence take-up, but not energy savings. For example, we find that high-income households are approximately twice as likely to participate in the program as low-income households. This suggests that low-income households in hot areas would appear to be a particularly valuable potential target. There is also a strong equity argument for targeting these programs to underserved groups.

In terms of policy recommendations, it might make sense to consider eliminating the program in mild climate zones. Our results indicate approximately zero average savings in mild areas (Climate Zones 6, 8, and 16), consistent with these units not being used enough hours of the year to generate substantial savings. Given scarce utility resources it makes sense to spend money where there is the biggest possible return on investment.

In addition, it would make sense to perform additional analyses of the program in warm climate zones. Savings are modest enough in these areas that it calls for a full-scale cost-benefit analysis. Critical for this broader analysis is the question of what fraction of participants are inframarginal "free riders" i.e. getting paid to do what they would have done anyway. The problem with inframarginal participants is that they add cost to the program without generating actual energy savings. This is an important consideration for all energy-efficiency programs (Boomhower and Davis, 2014), but particularly when the average savings are relatively modest.

Potentially offsetting this concern about inframarginal participants is the fact that this program tends disproportionately to deliver energy savings during high-value hours. We found in the second phase of the analysis (Boomhower and Davis, 2016), that this significantly increases the overall value of the program, especially once we account for the large capacity payments received by generators to guarantee their availability during high-demand hours. In particular, we find that the program has a timing premium of 50%, in that it delivers savings that are 50% more valuable than under a naive calculation that ignores timing.

More broadly, our work highlights the enormous potential of smart-meter data. Our econometric analysis would have been impossible just a few years ago with traditional monthly billing data, but today more than 50 million smart meters have deployed in the United States alone. This flood of high-frequency data can facilitate smarter, more evidence-based energy-efficiency policies that are better integrated with market priorities. California can help lead the way on this type of smart-meter analysis, both because of the innovative approaches through which California utilities make data available to researchers, and because of the California Energy Commission's long-standing support for evidence-based policy making.

Table 1: Take-Up and Energy Savings, Paired Regressions

	(1) Number of Households	(2) Takeup Rate (%)	(3) Average Savings per Replacement (kWh/year)
A. Baseline Specificiation			
All Households	7,284	0.13	398.6
			(34.2)
B. By Annual Household Income in Zip Code			
Below $75,000 Median	3,786	0.11	508.9
			(48.2)
$75,000+ Median	3,498	0.19	284.8
			(48.8)
C. By Educational Attainment in Zip Code			
Less than 30% with College Degree	3,724	0.12	324.1
			(45.1)
More than 30% with College Degree	3,560	0.16	488.1
			(51.8)
D. By Climate Zone			
Mild Areas (Zones 6, 8, & 16)	829	0.03	-29.1
			(82.9)
Warm Areas (Zones 9 & 10)	4,954	0.23	267.4
			(38.1)
Hot Areas (Zones 13, 14, & 15)	1,501	0.23	1,199.0
			(103.6)
E. By Racial Makeup of Zip Code			
Less than 50% Non-White	4,077	0.20	542.8
			(47.8)
More than 50% Non-White	3,207	0.09	228.5
			(48.5)
F. By Average Household Size in Zip Code			
Less than 3 People per Household	3,697	0.14	593.9
			(52.2)
More than 3 People per Household	3,587	0.13	222.2
			(44.5)

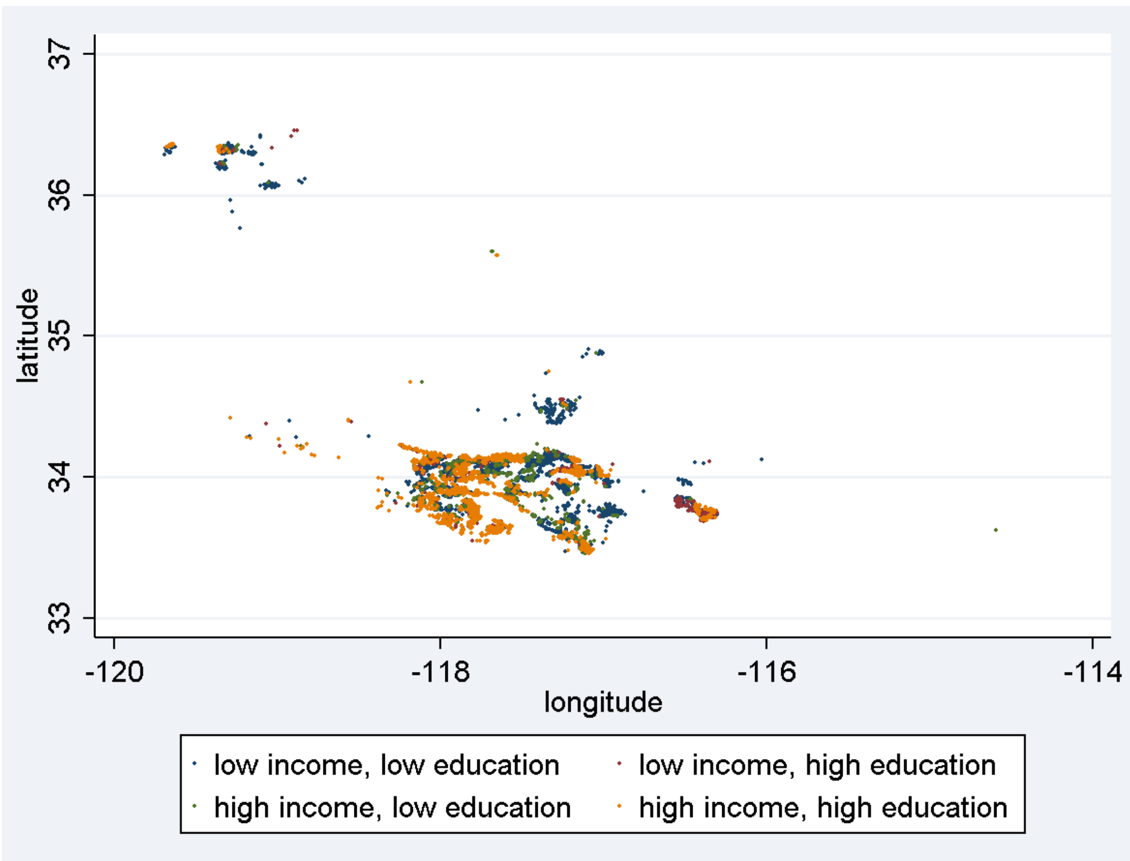
This table describes takeup and energy savings by zip code-level demographics. Takeup rates in Column (2) are percentages of SCE residential customers. Energy savings estimates and standard errors in Column (3) are estimated with separate regressions in each row. The dependent variable is average hourly electricity consumption at the household by week-of-sample by hour-of-day level. Average savings is the weighted sum of the coefficients on 288 indicator variables measuring the effects of replacement by month-of–year and hour-of-day. The weights are the number of days in the month. All regressions include household by hour-of-day by month-of-year, and week-of-sample by climate zone fixed effects. The 8 weeks prior to the replacement date are excluded. Standard errors are clustered at the household level.

Table 2: Take-Up and Energy Savings, Single Regression

	Average Annual Energy Savings (kWh/year)	
1[New Air Conditioner]	1.3	(16.3)
$\times$ Median Income > \$75,000	-65.8	(10.5)
$\times$ College Completion > 30%	166.3	(10.4)
$\times$ Warm Climate Zone	378.0	(14.4)
$\times$ Hot Climate Zone	1,166.3	(21.4)
$\times$ > 50% Non-White	-190.6	(9.4)
$\times$ > 3 People per Household	-68.4	(9.8)

This table reports coefficient estimates and standard errors from a single least squares regression. As with Table 1, the dependent variable is average hourly electricity consumption and the regression includes household by hour-of-day by month-of-year, and week-of-sample by climate zone fixed effects. The difference is that the specification includes a set of interaction between $1[NewAirConditioner]$ and the different heteogeneous factors listed in the row headings. Due to computational constraints, the reported standard errors above are only approximate and were calculated assuming that the covariances are zero between all hour-of-day and month-of-year cells.

Figure 1: Quality Installation Program Participants, By Income and Education



References

Boomhower, Judson and Lucas W Davis, "A Credible Approach for Measuring Inframarginal Participation in Energy Efficiency Programs," *Journal of Public Economics*, 2014, *113*, 67–79.

_ **and** _ , "Do Energy Efficiency Investments Deliver at the Right Time?," *Energy Institute at Haas Working Paper Number 271*, 2016.

Borenstein, Severin and Lucas W Davis, "The Distributional Effects of US Clean Energy Tax Credits," in "Tax Policy and the Economy, Volume 30," University of Chicago Press, 2015.