Impact Evaluation of Smart Thermostats

Residential Sector – Program Year 2018 Update

EM&V Group A

CALIFORNIA PUBLIC UTILITIES COMMISSION

CALMAC ID: CPU0205.03

December 22, 2020

DNV GL Energy Insights USA, Inc.

Table of Contents

List of Tables

List of Figures

1 BACKGROUND

In program year 2018 (PY2018), California's investor-owned utilities (IOUs) offered smart thermostats to their customers through energy efficiency rebate and direct install programs. Approximately 200,000 customers received smart thermostats under these programs. Rebated smart thermostats represented about 25%-30% of the total thermostats distributed and the majority were installed in single family homes. In most cases, the rebated thermostats were installed as the only energy efficiency measure enabling the attribution of energy consumption change post installation to the effect of these devices.

DNV GL conducted an evaluation of smart thermostats using data from rebate programs. Findings from this analysis were provided for the PY2018 evaluation cycle bus-stop (the "March Report").[1] The March Report found that energy savings per unit, in the form of cooling and heating energy consumption reduction, were less than 50% of claimed (expected) unit savings.

Since smart thermostats were offered on an opt-in basis, DNV GL based its evaluation on a quasi-experimental design using a matched comparison group. As discussed in the March Report, quasi-experimental designs have the potential to be affected by self-selection bias. In fact, the results from the March Report included a special adjustment to address evidence of self-selection.

This follow-on study was undertaken to investigate alternative approaches that could address self-selection bias more directly and effectively. The intent was to either provide additional support for the existing results or offer better results if alternative approaches are justified.

2 MOTIVATION

The main motivation for revisiting the PY2018 smart thermostat evaluation was to assess further the hypothesis that self-selection bias confounds smart thermostat savings estimation. Smart thermostat manufacturers have pointed to the demographic make-up of their adopters as younger and more affluent than the general population and more likely to increase consumption on average, year over year, than non-adopters.[2] For example, smart thermostat adopters may be more likely to have a child during the evaluation period, which results in a parent staying home with an associated increase in energy consumption. Smart thermostat adoption may also be correlated with a higher propensity to adopt an electric vehicle (EV), which would also increase consumption in the post-installation period. In all these instances, evaluation approaches could underestimate savings.

In practice, the only feasible way to evaluate a smart thermostat program is using a matched comparison group. The potential for self-selection bias is a known limitation of the approach. While this approach can provide a comparison group that reflects current characteristics of participants, it has limited ability to select comparison group households that reflect the underlying consumption trends that participants would have had if they had not taken part in the program. Lacking direct knowledge of changes to household composition due to babies or the adoption of EVs, for example, it is not possible to control for their possible differential presence explicitly.

In the March Report we found the following evidence that supported a hypothesis of differential energy consumption trends between participants and their matched comparison counterparts:

[1] http://www.calmac.org/publications/CPUC_Group_A_Report_Smart_Thermostat_PY_2018_CALMAC.pdf

[2] This hypothesis regarding participants being younger and more affluent is corroborated by survey findings summarized in the March Report.

- Difference-in-difference estimates indicated an increase in baseload among participants. As smart thermostats are not generally expected to affect baseload, this could be a manifestation of the differential trend.
- Self-reported survey results indicated that participants undertook more load building actions than non-participants they were matched to; for instance, they were more likely to purchase an EV than non-participants.
- Average daily energy values (particularly of electricity) indicated slightly higher use over time, in the pre-installation period, for participants than non-participants.

The March Report also offered alternative explanations for some of this evidence. For example, increases in baseload consumption among participants could reflect increase in usage of ventilation fans facilitated by the scheduling features of smart thermostats. Ultimately, the savings estimates were adjusted to reduce the "bias" indicated by the increase in baseload noted for participants. This approach assumed all the increase in the baseload was due to bias and that heating and cooling were downwardly biased on the same percentage basis. Despite the adjustment, the estimated savings were generally less than 50% of the expected savings.

In this report, we take two different approaches to further investigate the issue of self-selection. We attempt to account for trend directly in the matching process by selecting a comparison group that shares the underlying trend of the participant group. We also take advantage of the additional lag time since the end of 2018 to develop a comparison group made up of subsequent participants.

Each of these methods has the potential to address at least part of the underlying conditions that spur a self-selection process. The use of trend in matching could bring information into the process that would reflect the likelihood of an increase or decrease in consumption. If the addition of trend in matching has a consistent directional effect, then it would further support the hypothesis that self-selection exists even if the method may not fully adjust for it.

Using future participants as eligible comparison group members addresses the issue by requiring comparison group members to share the key attribute of having claimed a rebate for a smart thermostat. As a result, in addition to having similar observable characteristics, comparison group members will also share unobservable characteristics of program participants. For example, current and future participants are likely to have a similar propensity to adopt EVs, so increasing consumption due to EVs should be, on average, the same for participants and a comparison group built with future participants.

In fact, the only limitation of a comparison group composed of future participants, with respect to controlling for self-selection, is if the increase in consumption or trend is directly or indirectly correlated with the specific installation timing of the thermostat itself. Practically, this would mean that participants are more likely to purchase an EV within 12 months on either side of the thermostat installation date than at other times. While this could still be the basis for self-selection bias in the estimates, it is part of a much narrower range of possibilities.[3]

It is not possible to conclusively prove whether self-selection has affected the savings estimates of smart thermostats. With this additional analysis it is possible to see whether savings estimates change appreciably under comparison groups with different levels of self-selection risk. The comparison of these additional results will make it possible be more confident whether the existing savings estimates, both the original

[3] Similarly, the installation of the thermostat would have to be specifically associated with the timing of a baby's arrival, again, either directly or indirectly through some other process that is correlated with both the addition of a baby and the thermostat.

unadjusted estimates and those adjusted to account for self-selection, are reasonable or need to be reconsidered.

3 SELF-SELECTION BIAS

We provide a high-level overview of our methods to address self-selection bias and differential energy use trend, which are manifestations of this bias, in Table 3-1 below. The table includes a summary of the methods considered, the rationale for each, variables used to operationalize them, and the effectiveness of each approach. We provide details on the methods used in Appendix A and outcomes based on them in Appendix B.

Table 3-1. Summary of methods and outcomes used to address self-selection bias

Method	Rationale	Variables used	Test of effectiveness	Test outcomes
Include one-year trend in matching	Form matched groups with similar energy use trends	• One-year baseload trend • One-year weather normalized energy use trend	Compare post-pre change in energy consumption changes of the top 50% and bottom 50% of test groups and their matches	Inconsistent improvements in selecting matching groups with similar energy use trends
Include trend proxies in matching	Form matched groups with similar energy use trends	• Tenure • Propensity-to-change	Compare post-pre change in energy consumption changes of the top 50% and bottom 50% of test groups and their matches	Slight improvement in selecting matched groups
Select data from future participants	Form matched groups with similar energy use trends	Information on energy use	Balance in matched samples	Well-balanced samples with respect to matching metrics

4 SAVINGS ESTIMATES

We provide updated savings estimates for smart thermostat installations among rebate program participants in this section based on the methods used to deal with self-selection outlined above. We also compare these results to savings provided in the March Report and expected electricity and gas savings based on submitted claims.

4.1 Future participant comparison group savings estimates

This analysis continues the evaluation of smart thermostat savings of PY2018 participants. Primary new results are savings estimates using a future (PY2019) participants' comparison group. This comparison group is formed by matching consumption and seasonal variables of customers who received their thermostat at least a year later than their 2018 counterparts. In addition to being matched on site-specific characteristics, future participants should share any unobservable characteristics associated with a decision to opt into such a program. The use of data from future participants is considered the best way to control for the effects of self-selection and the potential differential trend in energy consumption noted among the participants that we evaluated.[4]

The results in Figure 4-1 (for electricity) and Figure 4-2 (for gas) provide smart thermostat savings estimates based on analyses that use future comparison groups (future comparison group results). Across all IOUs and fuels, the future comparison group results (Future, yellow) continue to be less than 50% of expected savings (Expected, dark blue). Additionally, the future comparison group results do not indicate increased savings relative to the final published March adjusted results (March Adjusted, light blue) on the

[4] We provide future participant counts used in the study in Appendix A. Appendix C includes a review of the number of current participants used in the analysis.

left of each bar graph. For the electric estimates, the future comparison group results fall between the base and adjusted estimates from March Report in every case. For gas, they are in that range or lower. Recognizing that the future participant results are less likely to be affected by self-selection, these results indicate that March adjustments were reasonable, if not even somewhat generous.

Figure 4-1. Electricity percent savings per smart thermostat for future participant comparison groups

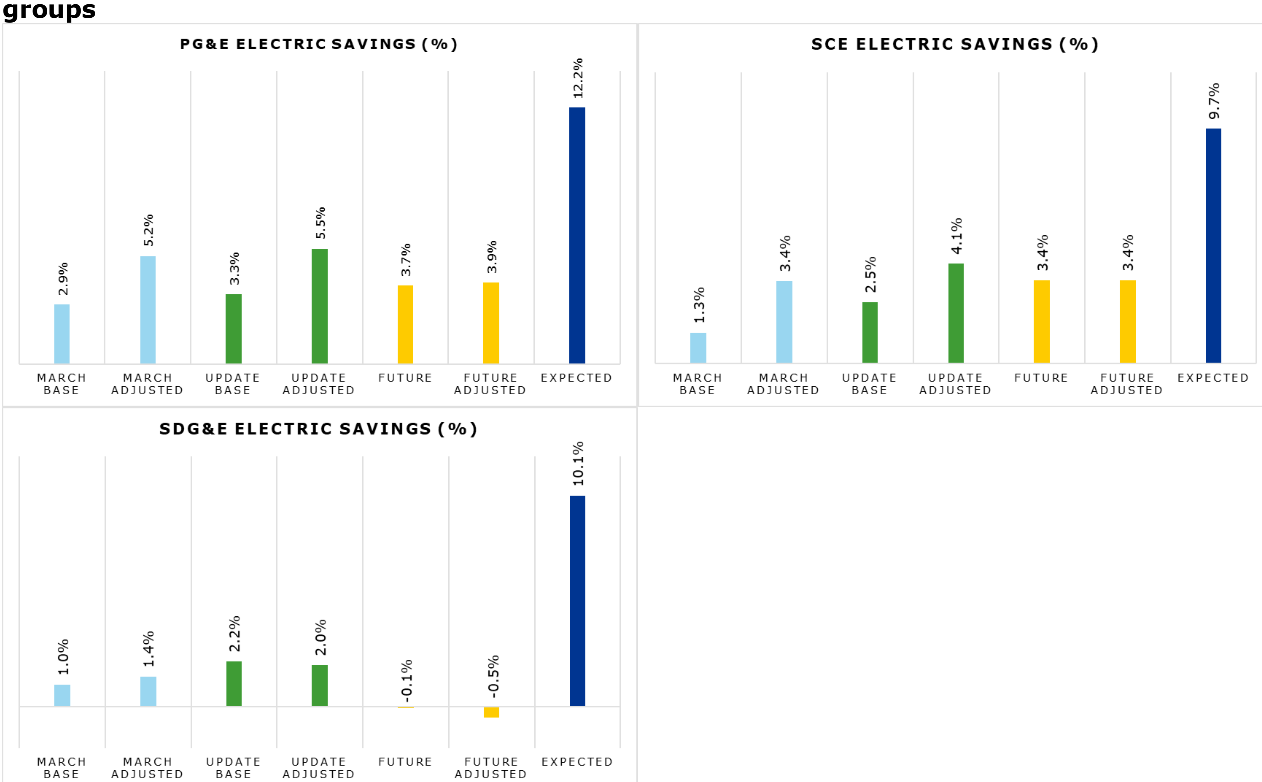
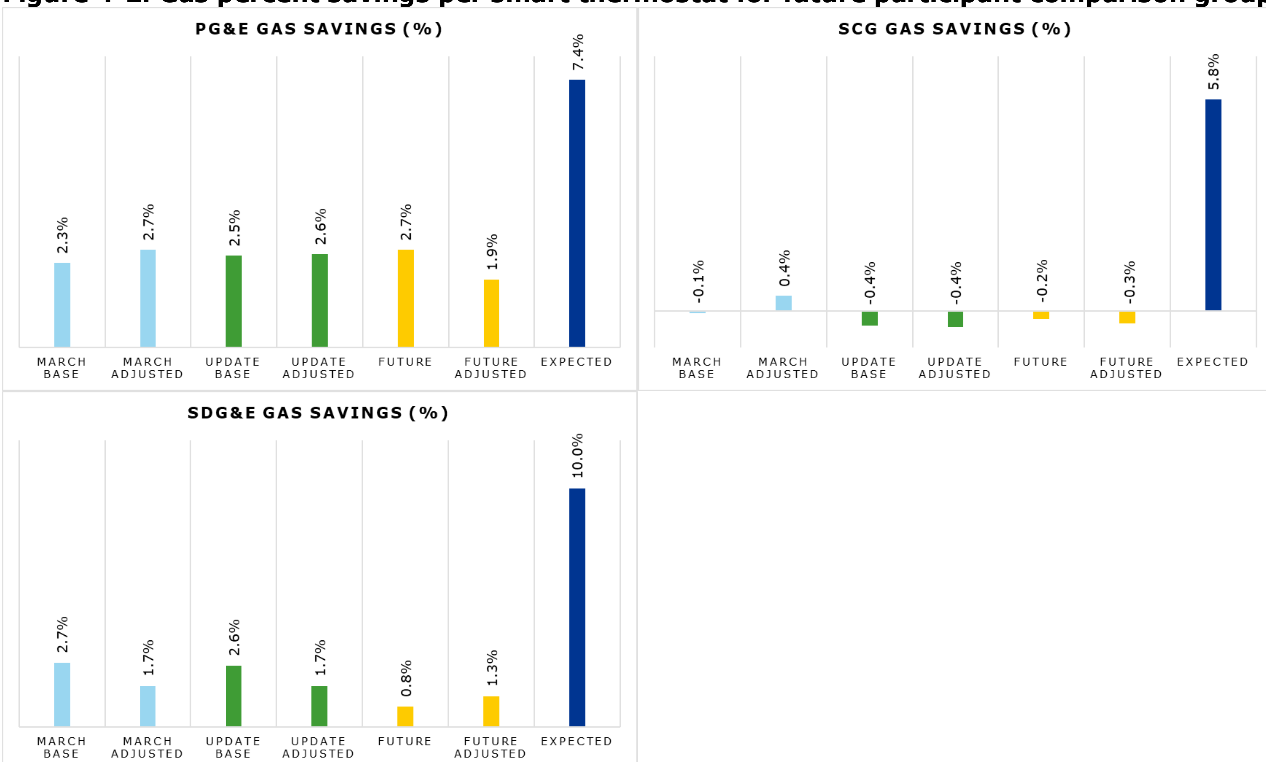


Figure 4-2. Gas percent savings per smart thermostat for future participant comparison groups



These conclusions are further supported by comparing the future comparison group results to updated base and adjusted results (in green). These results update the March findings using new data, consistent matching methods and, most importantly, a matching window immediately prior to the implementation period, which is consistent with the way future participants are matched.[5]

The updated results put the March results on an equal footing with the future participants results in every respect. For the electric results, the updates increase both base and adjusted results modestly. Again, future participant results fall between the base and adjusted updated results, indicating adjustments were needed, but at a lower level. The gas future participant results are reasonably consistent with prior results except for SDG&E, where smaller populations lead to greater volatility in savings estimates.

Finally, Figure 4-1 and Figure 4-2 include one more important set of results. The adjustments applied in the March Report and updated results account for an increase in baseload consumption by participants that could be a manifestation of self-selection. We applied the same adjustment approach to the future participant estimates to see if there was any need for an adjustment. The adjustments to the future comparison group results are negligible or negative. This is further evidence that the future participant results address concerns related to self-selection.

4.2 Trend-proxy comparison group savings estimates

A secondary effort from this analysis attempted to address potential differential trend through the inclusion of trend-proxy variables when selecting comparison groups. Attempts to identify trend in the pre-installation

[5] The updated results are based on new initial 10:1 matches that reflect an improved matching approach compared to the one we used for the March Report. Unlike the March Report, which used initial matches based on monthly billing data and propensity score matching (PSM) and final matches that were selected with replacement, the current approach matches households based on starting energy-use levels and seasonality based on distance-based methods.

data (one-year trend) as an indicator of pre-to-post installation trend failed in testing and we do not produce actual savings results for these efforts. What we referred to as trend-proxy approaches, using available data that are correlated with post-pre energy consumption change, appeared to offer promise in testing. Savings estimates across these efforts are presented here.

Figure 4-3 and Figure 4-4 provide percent electricity and gas savings based on participant and matched comparison group data obtained from matching that include trend proxies. The trend proxies used are tenure and propensity-to-change. Tenure is the length of time, measured in years, that a customer has resided at a premise. Propensity-to-change captures the probability of having an increasing trend in energy consumption as a function of tenure. Further discussion of the process to identify these relevant variables can be found in Appendix A and Appendix B.

It is evident that these variables do not change estimated savings over the base case. Despite tenure being shown to move the needle to some degree when matching for trend, this trend proxy does not change estimated savings in a consistent manner. Savings estimates based on data from trend-proxy matching were also adjusted to correct for the possible presence of self-selection bias. Unlike the case for estimates based on future participants, these adjusted estimates are higher than those based on the trend proxies alone, further corroborating that these trend proxies do not fully correct for self-selection bias.

Figure 4-3. Electricity percent savings per smart thermostat for trend-proxy comparison groups

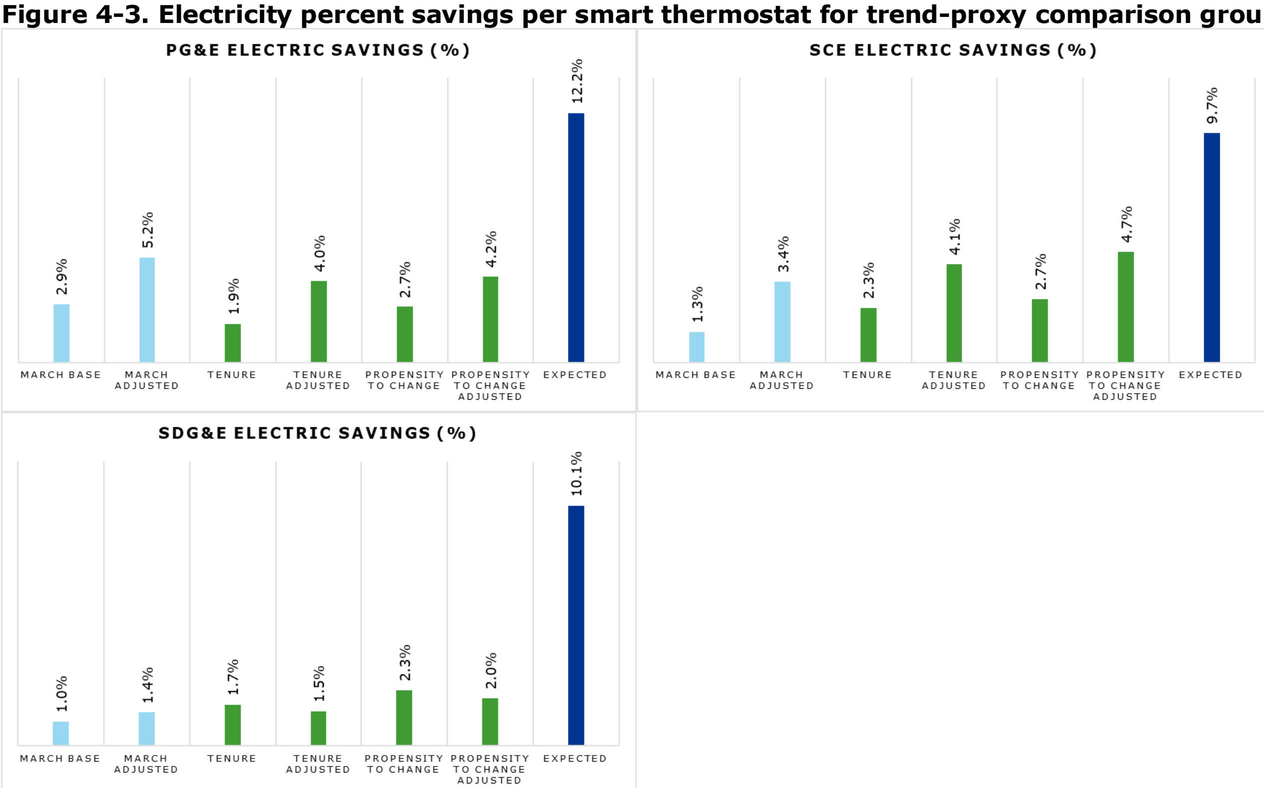
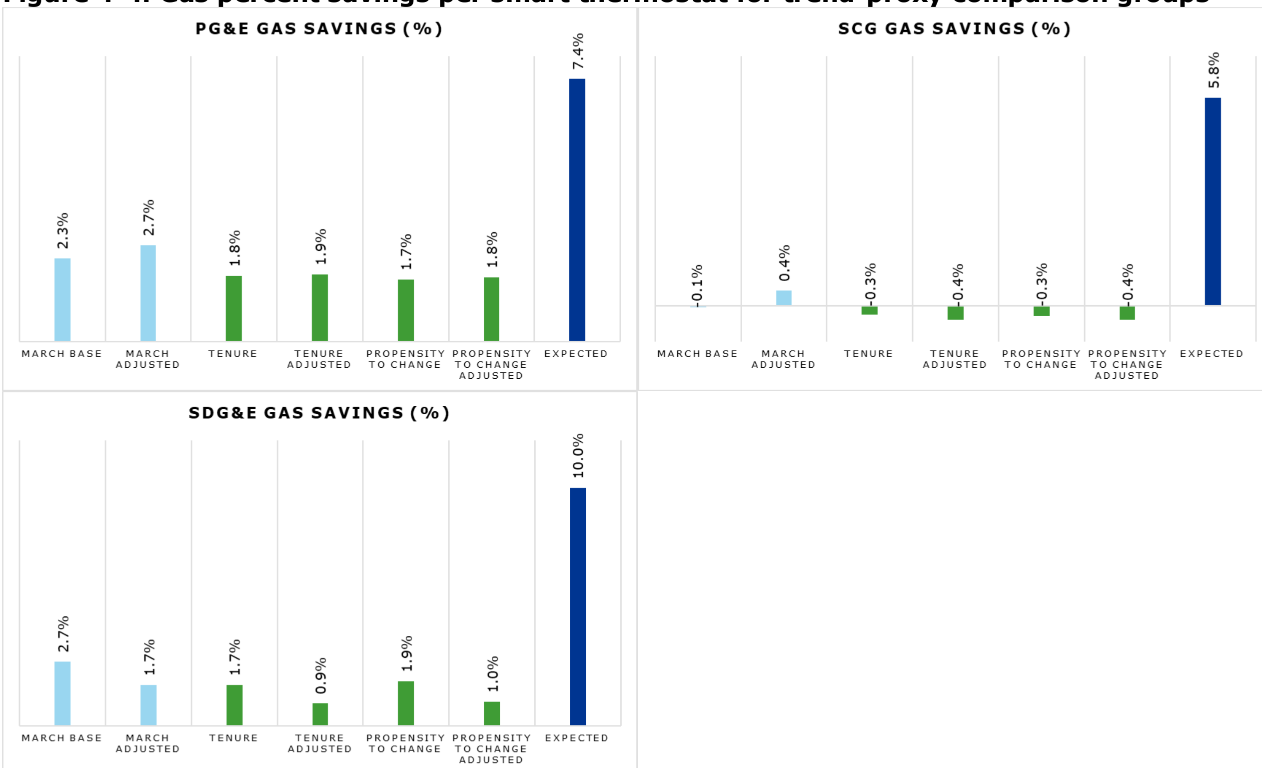


Figure 4-4. Gas percent savings per smart thermostat for trend-proxy comparison groups



5 CONCLUSIONS

This analysis confirms that future participant comparison groups offer the best feasible basis for constructing a comparison group. On a theoretical basis, future participant comparison groups address most concerns regarding potential self-selection as matched customers share those unobservable characteristics associated with opting to get a rebate for a smart thermostat. In practice, future participant comparison group results are roughly consistent with the results from the March Report. Perhaps most importantly, the future participant comparison group results stand on their own, not needing an adjustment.

While future participant comparison group results may be the strongest option, they may not always be feasible for a timely evaluation given the required time delay. In this respect, they provide useful feedback on the adjustment that was applied for the March results. The future participant comparison group results are consistently smaller in magnitude than the final adjusted results from March, but in cases where substantial upward adjustments were made, results generally support the need for some adjustment. The future comparison group results provide additional credence to an adjustment based on baseload increase but at some de-rated level.

The remaining results in this analysis illustrate the challenge of incorporating trend into matched comparison groups when future participant comparison groups are not available. Despite substantial effort across a variety of different strategies, we were unable to demonstrate any ability to better select matched comparison groups with respect to trend. Though there remain more opportunities for exploration in this area, for now, evaluations that do not use future participants will continue to need to consider at least a partial adjustment to address the effects of self-selection.

We will use the data and findings from this study to produce updated climate-zone level savings for customers with rebated smart thermostat for future workpapers.

6 APPENDIX

We provide details on the methodology used for the update study including methods we considered to correct for self-selection bias (Appendix A), outcomes based on these efforts (Appendix B), and the savings estimation methods we used (Appendix C). We also provide outcomes from the matching which underpins the savings estimates that inform our conclusions (Appendix D).

6.1 Appendix A: Approaches to Address Self-Selection Bias

In this section, we describe the methods DNV GL used to address the effect of possible differential trends in energy use (which are manifestations of self-selection bias) between smart thermostat program participants and selected non-participants whose data is used to inform the baseline in the evaluation. The methods involve the use of two different types of comparison groups that inform the baseline against which energy use changes are evaluated. The first is a comparison group matched with participants using initial energy usage levels and seasonality, and trend in energy use. The second group is future participants that self-select into smart thermostat rebate programs, whose data is used for this type of comparison. We describe the approaches including the data used in both cases below.

6.1.1 Trend in Matching

DNV GL undertook two approaches to include trend in energy use when matching participants to non-participants to identify not only households with similar usage patterns, but also trends. The first approach relied on identifying the one-year trend in energy use prior to installation that could be used in matching. This approach will be useful only if we are successful in identifying a trend with a single year of data, after controlling for weather sensitivity of energy use, and if that trend can predict trend in energy consumption in subsequent years.

The second approach relied on identifying household characteristics that are correlated with energy use trend and can be used to identify matched comparison groups. For this approach to work, we need to identify customer characteristics data that are correlated with consumption change over time, either in non-participant populations or participants during periods of non-participation, which can then be used as a basis for matching.

6.1.1.1 One-Year Trend

Measuring one-year trend

DNV GL estimated different one-year trends in energy consumption based on PRISM-models used to weather normalize electricity and gas consumption at the individual site level. Weather normalization makes it possible to determine trends in energy use based on typical or normal weather, effectively removing the impact of yearly weather fluctuations on energy use. The method involved estimating a set of regression models of energy use as a function of weather. We used three different site-level modeling approaches to estimate the presence of trend:

1. **Baseload Trend:** We determined baseload-only days, which are days with no cooling or heating degree-days indicating the absence of weather correlated energy use, using a PRISM model without trend. Baseload-only days were then re-estimated with a trend term to estimate trend in energy use on such days.
2. **PRISM with Trend Added:** We used the basis points identified as part of the PRISM model's grid search, without trend, to re-estimate models with the optimal basis points and a trend variable. This approach for identifying trend includes a trend variable in the PRISM model after the grid search.

3. **Trend Optimized:** This modeling approach for identifying trend includes a trend variable as part of a PRISM model's grid search. A trend term is estimated along with optimal degree days for each site in this model.

The three PRISM-based trend models used to determine one-year trend estimates are specified in the following way:

$$E_{im} = \beta_0 + \beta_h H_{im}(\tau_h) + \beta_c C_{im}(\tau_c) + \beta_T T_i + \varepsilon_{im}$$

Where:

E_{im} - Average electric (or gas) consumption per day for participant i during period *m*.
$H_{im}(\tau_h)$ - Heating degree-days (HDD) that participant i faces during period *m* at the heating reference temperature, τ_h.
$C_{im}(\tau_c)$ - Cooling degree-days (CDD) that participant i faces during period *m* at the cooling reference temperature, τ_c, (not included in gas models).
T_i – Trend variable for participant i.
$\beta_0, \beta_h, \beta_c, \beta_T$ – Site-level regression coefficients measuring intercept (base load), heating load, cooling load, and trend for a single year's energy consumption, respectively.
τ_h - Heating reference temperatures, determined by choice of the optimal regression model.
τ_c - Cooling reference temperatures, determined by choice of the optimal regression model.
ε_{im} – Regression residual.

Additional information about how reference temperature points are selected and how we estimated weather normalized consumption is included in Appendix C.

Measuring effectiveness of one-year trend

To assess the effectiveness of the three one-year trend variables in matching, we used post- to pre-installation period energy consumption changes ("post-pre change") as a measure of energy consumption trend and, thus, as the metric used to gauge similarity in such trends. Positive values for this metric indicate increasing trend while negative values indicate decreasing trend. We included the one-year trend variables along with the 'base-case' or status-quo matching covariates (total energy use and the ratio of summer to winter energy use) to test their effectiveness in matching households with similar energy use trend.

We used electricity data from the March PG&E and SDG&E matched comparison households to conduct a series of matching tests. The tests involved randomly splitting the comparison group households into test (pseudo-treatment) and comparison pools and matching the test group pool to households in the comparison pool. Table 6-1 presents the number of households selected for the exercise from PG&E and SDG&E. In the table, the column "Test Group" provides counts of households selected to be in test groups (where test group is 1) and comparison pools (where test group is 0).

Table 6-1. One-year trend matching test groups

IOU	Test Group	Customer Numbers
PG&E Electric	0	3,948
	1	980
SDG&E Electric	0	3,798
	1	943

Particularly, we selected 20% of customers randomly to form three different test groups, each of which was matched to the remaining 80% using matching approaches with and without the one-year trend measures. We tested the success of each matching strategy by comparing how closely post-pre changes for the selected comparison groups resembled the top 50% and bottom 50% of these changes for the test groups. A successful matching approach with one-year trend would be expected to improve this resemblance over the base-case or status quo approach that uses only total annual energy consumption and the summer-winter ratio in matching.

6.1.1.2 Trend Proxy

Establishing trend proxies

As an alternative, DNV GL also considered incorporating household characteristics that are good proxies for energy use trend in the matching approach. Prominent among the characteristics considered was the "age" of the household, partly precipitated by comments from smart thermostat manufacturers that "young" households with growing families are likely to have higher energy use trends over a specified period than "older" households. Based on available data, we focused on tenure or the length of time, measured in years, a customer has resided at a premise as a possible proxy for household age and, thereby, for energy use trend.

DNV GL first sought to establish if this variable was a good proxy for household age using data from the non-participant survey conducted for the March Report. We explored correlations between tenure and household characteristics including the number of children under five years old and the number of adults over 65 years in a home for this purpose. After investigating the usefulness of tenure as a good proxy for age, we also tested how well correlated it is with post-pre change (energy consumption trend). Both energy use data and data from the non-participant survey were used to examine the usefulness of tenure as a proxy for energy use trend.

DNV GL also established an additional proxy for energy use trend that we call "propensity-to-change" based on the relationship between post-pre energy consumption change (delta) and tenure. To accomplish this, we estimated linear models of post-pre delta values as a function of tenure. These models were estimated using data from matched comparison households that do not face any intervention which changes their energy use in a systematic manner. Details of the data used for these estimates are provided in Appendix D.

The variability of post-pre deltas is large, and thus, models based on tenure are not likely to explain such variability fully. However, fitted values from these models can serve as an additional proxy for energy use trend that can be used in matching. DNV GL estimated the following regression model to obtain such fitted values:[6]

$$delta_i = \alpha_i + \beta_{i1} * tenure + \beta_{i2} * tenure^2 + \beta_{i1} * tenure^3 + \epsilon_i$$

Both tenure and estimates of propensity-to-change were included in variants of the base-case matching model one at a time to test their effectiveness in matching.[7]

Measuring effectiveness of trend proxies

DNV GL used data from about 60,000 PG&E electric comparison households based on new 10:1 matches undertaken in the of summer 2020 to develop and test the effectiveness of trend proxies in matching.[8] The

---

[6] An investigation of average post-pre delta by tenure using electricity data from PG&E, SDG&E, and SCE's comparison group households indicated the relationship between the two variables follow a quadratic or cubic function. DNV GL found the cubic function to fit the data well with estimated coefficients that are similar and significant across different runs.

[7] The base-case variant matching model uses annual energy, the summer-winter ratio of energy use, and, for electricity, 6 p.m. kWh on identified heat wave summer days, as matching variables. Details on how the 6 p.m. kWh variable is constructed is provided in the data section in Appendix C.

[8] The new matches were based on annual energy and summer-winter ratios and Mahalanobis distance matching. Details of the data used in matching is provided in Appendix C.

tests were based on simulations that involved drawing 20% of households randomly to form a test group, conducting matching, storing post-pre energy use changes for matched households, and replicating this process 1,000 times.

We tested the success of matching strategies that involve trend proxies by comparing:

- The mean and median of the top 50% simulated post-pre changes of test groups to mean post-pre changes of their matches, and the bottom 50% simulated post-pre changes of test groups to mean post-pre changes of their matches
- Histograms of the percent difference between the top 50% simulated post-pre changes of test groups and post-pre changes of their matches, and the bottom 50% simulated post-pre changes of test groups and post-pre changes of their matches

A successful 'proxy-trend enhanced' matching approach would be expected to move comparison groups' trend closer to those of test groups compared to the status quo approach that uses only annual energy use, summer-winter ratios, and 6 p.m. demand for matching.

6.1.2 Future Participants

We also used data from future (PY2019) rebate program participants to compare the effect of smart thermostats on energy use. Data from future participants is suitable to form comparison groups since the self-selection mechanism that drives current participants is likely to be present among future participants whose energy use trends will, thus likely be similar to those of current participants.

We received daily data for future participants covering the period January 2017 to December 2019 for use in the analysis. Table 6-2 provides counts of future participants by IOU and the final count used in the analysis. Only data from future participants who installed smart thermostats in 2019, do not have on-site PV, and have adequate pre- and post-installation data prior to their own 2019 installation dates are used in the analysis. Energy consumption data from future participants up to the period where they install smart thermostats is used in the analysis since post-installation energy consumption for these households reflect the effect of smart thermostats.

Table 6-2. Count of future (PY2019) participants by IOU

IOU	Fuel	Count of PY2019 Claims	Count of PY2019 Participants in analysis
PG&E	Electric	13,772	4,440
	Gas	13,772	5,689
SDG&E	Electric	6,860	1,984
	Gas	6,860	4,357
SCE	Electric	3,488	1,423
SCG	Gas	14,762	5,983

We matched current (PY2018) to future participants using the base-case matching approach described above, which involves matching based on annual energy use, summer-winter energy use ratios, and for electricity, 6 p.m. kWh on identified heat wave summer days. Since the number of future participants is generally lower than current (PY2018) smart thermostat installers, we created installation cohorts and matched with-replacement within cohorts.

A future participant that has the requisite 12 months of pre- and 12 months of post-period data for analysis for multiple installation periods or cohorts can serve as a matching candidate to participants in each of those cohorts or installation periods. As a result, a future participant household can be selected multiple times both within a cohort and across a cohort during matching. The implications of this duplication on standard errors is addressed in the second stage model, where the precision of savings estimates is based on clustered standard errors.

6.2 Appendix B: Results of Self-Selection Bias Corrections

We discuss outcomes from efforts used to address self-selection bias in this section. First, we provide the results from the tests in matching based on one-year trend variables. Second, we provide outcomes in matching that include trend proxies. We describe test of balance from matching treatment and future participants households in Appendix D.

6.2.1 One-Year Trend

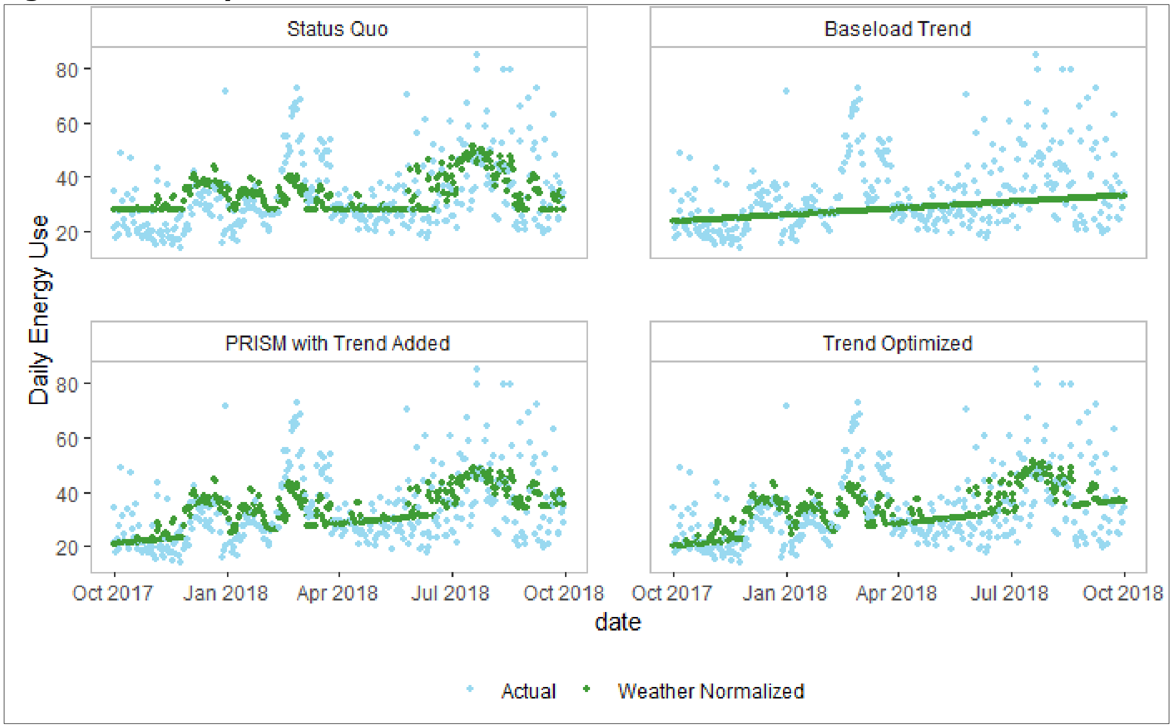
To the extent that these one-year trend estimates predict year-over-year change in energy consumption, they can be use useful in matching participant households to non-participants with similar changes thereby mitigating self-selection bias. As indicated in section 6.1.1, we tested the effectiveness of one-year trend terms in accomplishing this goal based on PG&E and SDG&E electricity data from the March comparison groups. We selected three test groups randomly from this data and matched them to the remaining comparison group households using the following two approaches:

1. The status-quo approach that matches households based on starting energy use and seasonality
2. The status quo model supplemented with one-year trend

Figure 6-1 presents an example of one-year trend fitted using the three approaches for a single site. To the extent that these one-year trend estimates predict year-over-year change in energy consumption, they can be use useful in matching participant households to non-participants with similar changes thereby mitigating self-selection bias. As indicated in section 6.1.1, we tested the effectiveness of one-year trend terms in accomplishing this goal based on PG&E and SDG&E electricity data from the March comparison groups. We selected three test groups randomly from this data and matched them to the remaining comparison group households using the following two approaches:

1. The status-quo approach that matches households based on starting energy use and seasonality
2. The status quo model supplemented with one-year trend

Figure 6-1. One-year trend fits from weather normalization models

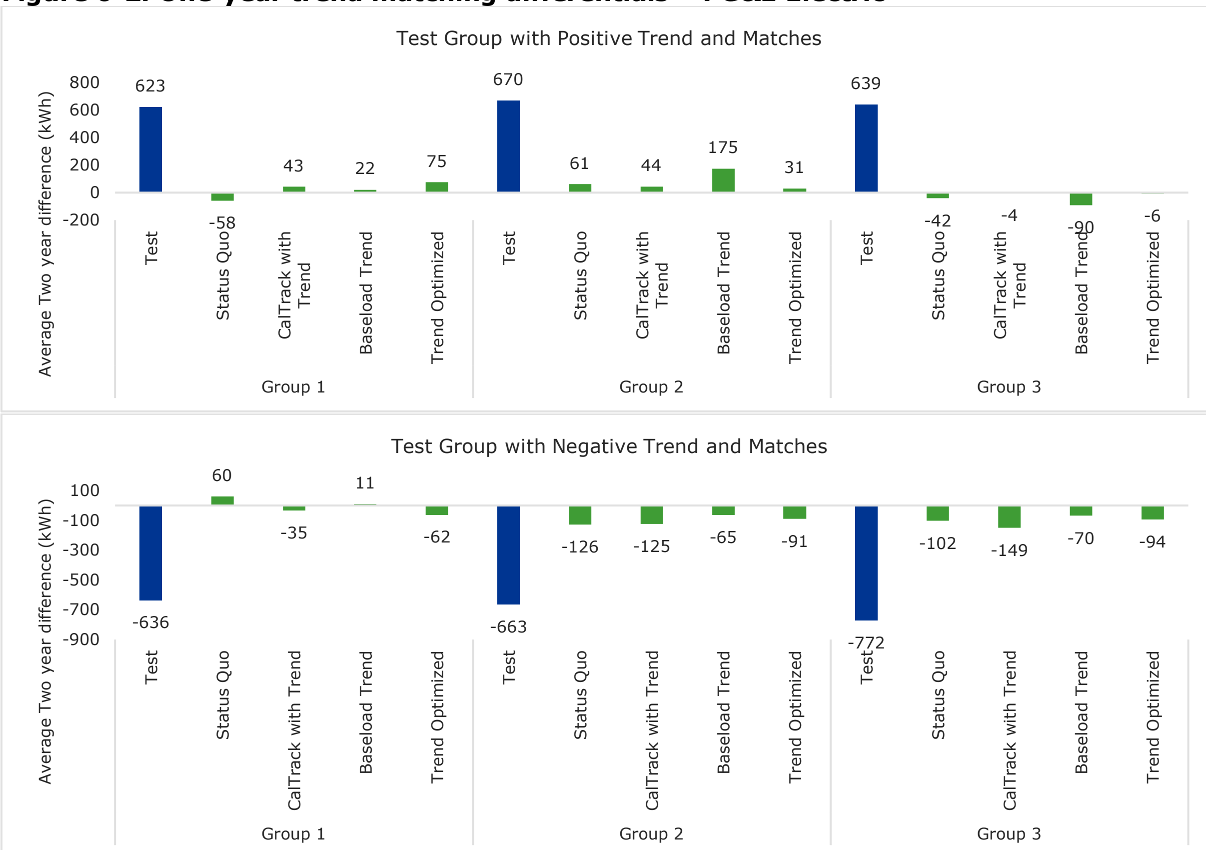


We evaluated whether including estimates of one-year trend improves matching outcomes by examining post-pre energy use changes of test groups and their selected matches. We used the top half post-pre consumption change of test groups as a measure of positive or increasing energy use trend and the bottom half as measuring negative or decreasing energy use trend. If matching with one-year trend provides improvement over the status quo approach, the average post-pre change of the matched comparison groups would be more like the test groups' than the average trend of comparison groups matched using the status quo approach.

Figure 6-2 provides results from this undertaking based on PG&E electric data; outcomes based on data from SDG&E are similar. For each group, the test value provides the average post-pre change for the test group, which is constructed to have substantial change in a positive or negative direction. On the other hand, post-pre changes of comparison groups matched using the status quo method are expected to be close to zero since there is nothing in the status quo matching process designed to address trend. The status quo results shown in the figure reflects such natural trends.

The three trend-based matched comparison groups, by contrast, were designed to reflect the trends in their associated test group matches. If the one-year trend based matching works, we would expect the average post-pre change of the selected groups to move consistently in the direction of the average post-pre changes of the test groups to whom they are matched. As indicated in the figure, there do not appear to be consistent improvements of this sort when including one-year trend variables in matching. For test group 1 with a positive trend, all three one-year trend matching variables appear to improve matches, however, these improvements do not persist for the other five replications. Overall, the results from these matching exercises indicate that one-year trend does not provide consistent improvement in matching over the status quo approach.

Figure 6-2. One-year trend matching differentials – PG&E Electric



6.2.2 Trend Proxy

We considered alternatives to the one-year trend to match households. We started with the hypothesis that the age of a household is related to energy use trend and can be proxied by tenure. Tenure measures the length of time a household has resided at a premise and is based on account start dates available in utility records. We measure tenure as the difference, in years, from such account start dates to the beginning of the analysis period for this study, which is 2017. This measure is rounded to the nearest integer such that households residing less than half year in their current homes relative to the start of 2017 are considered to have tenure of 0.

We examined how well tenure proxied the "age of households" by looking at correlation of tenure with demographic information, including the age composition of households, using data from the non-participant survey we conducted for the March Report. While the demographic information is limited to only survey respondents, it does give us valuable insight into whether tenure is a good proxy for household age and what other demographic features it could be associated with that may be driving differences in energy use over time.

Table 6-3 presents correlations between tenure and demographic variables for three IOUs based on data from the non-participant survey. We considered households with children under five years as young and those with adults over the age of 65 as older and examined how well tenure correlated with such households. We found that households with children under five years were not consistently associated with lower tenure while those with adults over the age of 65 were consistently associated with higher tenure. On the other hand, households that reported taking more energy use increasing actions had consistently shorter tenures.

Table 6-3. Tenure correlations with non-participant survey demographic variables

Variable	PG&E (n=309)		SCE (n=246)		SDG&E (n=261)	
	Correlation	P-value	Correlation	P-value	Correlation	P-value
Households with children under 5 years	0.05	0.43	-0.08	0.19	-0.19	0.00
Households with adults over 65 years	0.47	0.00	0.22	0.00	0.17	0.01
Number of energy use increasing actions	-0.20	0.00	-0.06	0.36	-0.16	0.01
Number of energy use decreasing actions	0.21	0.00	-0.04	0.56	0.13	0.04
Net energy use increasing actions[9]	-0.31	0.00	-0.01	0.94	-0.22	0.00

We also examined how trend in energy use (post-pre change) is associated with tenure and demographic information. As Table 6-4 shows, households with children under five years were not consistently associated with higher trend. However, households that had reported taking more energy use increasing actions tended to have higher energy use trend and shorter tenure (Table 6-3). Thus, it is not necessarily household age but actions that *households new to their homes* take that is associated with positive energy use trend.

Table 6-4. Trend correlations with non-participant survey demographic variables

Variable	PG&E (n=309)		SCE (n=246)		SDG&E (n=261)	
	Correlation	P-value	Correlation	P-value	Correlation	P-value
Tenure	-0.27	0.00	-0.02	0.73	-0.07	0.26
Households with children under 5 years	0.03	0.59	-0.20	0.00	-0.02	0.71
Households with adults over 65 years	-0.10	0.07	0.00	0.95	-0.11	0.09
Number of energy use increasing actions	0.31	0.00	0.20	0.00	-0.05	0.39
Number of energy use decreasing actions	-0.15	0.01	-0.04	0.51	-0.07	0.25
Net energy use increasing actions[10]	0.31	0.00	0.19	0.00	0.02	0.75

Further, using consumption data from comparison group homes from the March Report, correlations between tenure and trend in energy use indicate that households with longer tenure tend to have lower trend (post-pre change) than households with shorter tenure. Since these households had no energy efficiency intervention through utility programs, all post-pre changes can be used to gauge the association of tenure with energy use trend. Table 6-5 provides these correlations. It indicates that decreases in post- to pre-installation percent change in energy (indicating declining energy use trend) is weakly but negatively correlated with tenure, indicating that as the age of a household increases, energy trend decreases.

Table 6-5. Correlation between percent in energy use change and tenure

Variable	Post-pre energy use change (trend in energy use)		
	Correlation	P-value	Number of Observations
PG&E tenure	-0.076	0.00	4,928
SDG&E tenure	-0.034	0.01	4,720
SCE tenure	-0.061	0.00	5,358

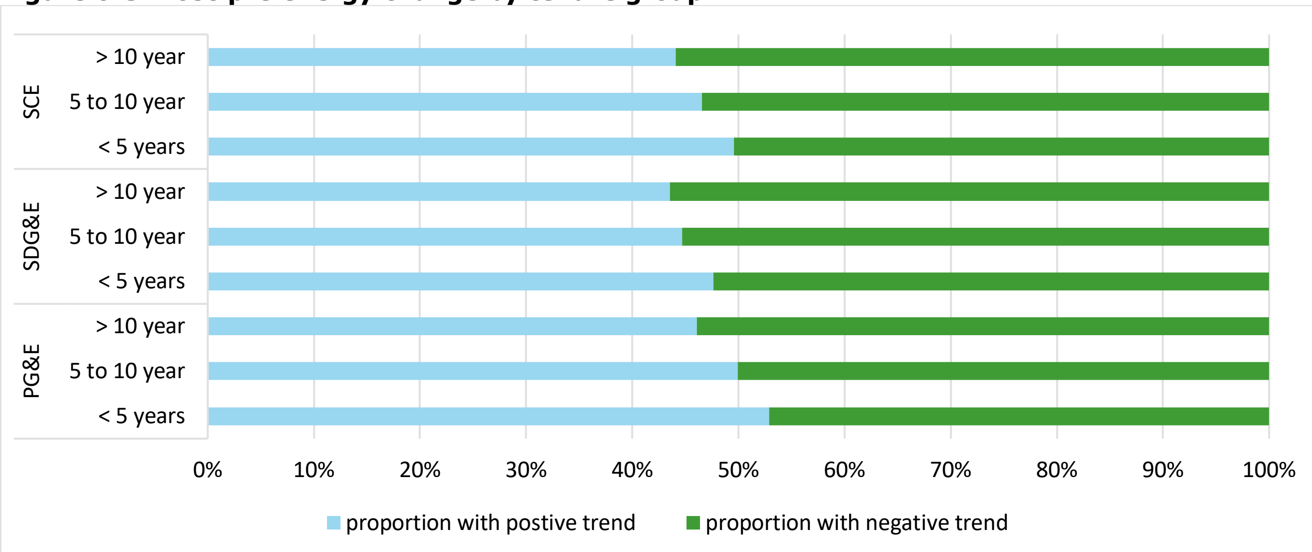
Further evidence of the negative association between tenure and trend (post-pre energy use change) can be found by looking at the proportion of positive and negative changes for this metric among different tenure cohorts using the same March Report comparison group data. Figure 6-3 provides the proportion of households with positive (increasing) and negative (decreasing) post-pre energy use change among comparison group households used in the March Report. It indicates that a larger proportion of households

[9] Net increase is derived as the difference in the proportion reporting an action that would increase energy use and the proportion that report doing the opposite which would result in decreased energy use for that action: Positive if net increase, negative if net decrease.

[10] Ibid.

at their current residence five years or less had positive energy use trend across all three IOUs. In general, the figure shows that households with lower tenure were more likely to have a positive trend than households with longer tenure.

Figure 6-3. Post-pre energy change by tenure group



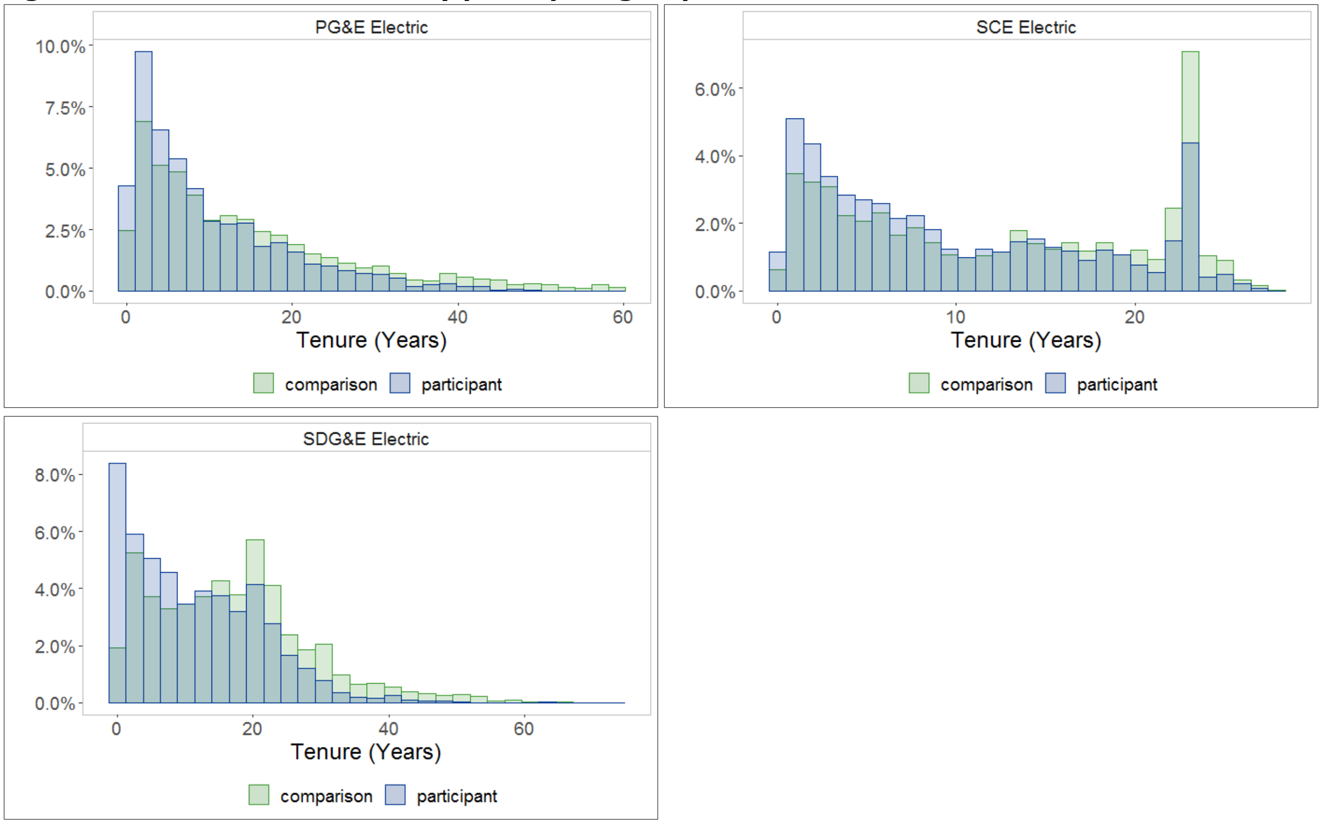
The negative association between tenure and trend is important in light of new evidence that shows differences in the distribution of tenure between participant and matched households used in the March Report. Table 6-6 provides the percent of participant and non-participant households in different tenure groups. It indicates that participants had shorter tenure than the non-participants to whom they were matched. For example, the percent of participants with a tenure of five years or less was substantially higher than non-participants to whom they were matched. On the other hand, the percent of non-participants with tenure of 10 years or greater was substantially higher than participants to whom they were matched.

Table 6-6. Percentage of participants and non-participants by tenure groups

IOU	Group	tenure < 5 years	tenure 5 to 10 years	tenure > 10 year
PG&E	Participant	40%	22%	37%
	Non-participant	28%	20%	52%
SDG&E	Participant	33%	17%	49%
	Non-participant	18%	14%	69%
SCE	Participant	37%	22%	41%
	Non-participant	28%	18%	54%

Figure 6-4 provides distributions of tenure for participants and matched non-participants featured in the March Report and indicate the same finding. The distributions indicate that more participants were represented in bins indicating shorter tenure than non-participants whereas more non-participants were in bins indicating longer tenure.

Figure 6-4. Distribution tenure by participant group



Given the evidence that tenure has an inverse correlation with trend and the imbalance in tenure between participants and non-participants, we conducted a series of tests to investigate the effect of including tenure and propensity-to-change, constructed as a function of tenure (see Appendix A), in matching. The tests to determine the effectiveness of tenure in matching homes with similar trends were based on data from PG&E electric comparison households used in the current study.

These tests involved running 1,000 matching simulations to test alternative matching strategies. Each simulation involved drawing 20% comparison group homes randomly as a test group and matching them to the remaining comparison group homes. The three matching strategies of interest were:

- Base case matching using total annual energy use, summer-winter ratio, and 6 p.m. kWh
- Base case matching variables and tenure
- Base case matching variables and propensity-to-change

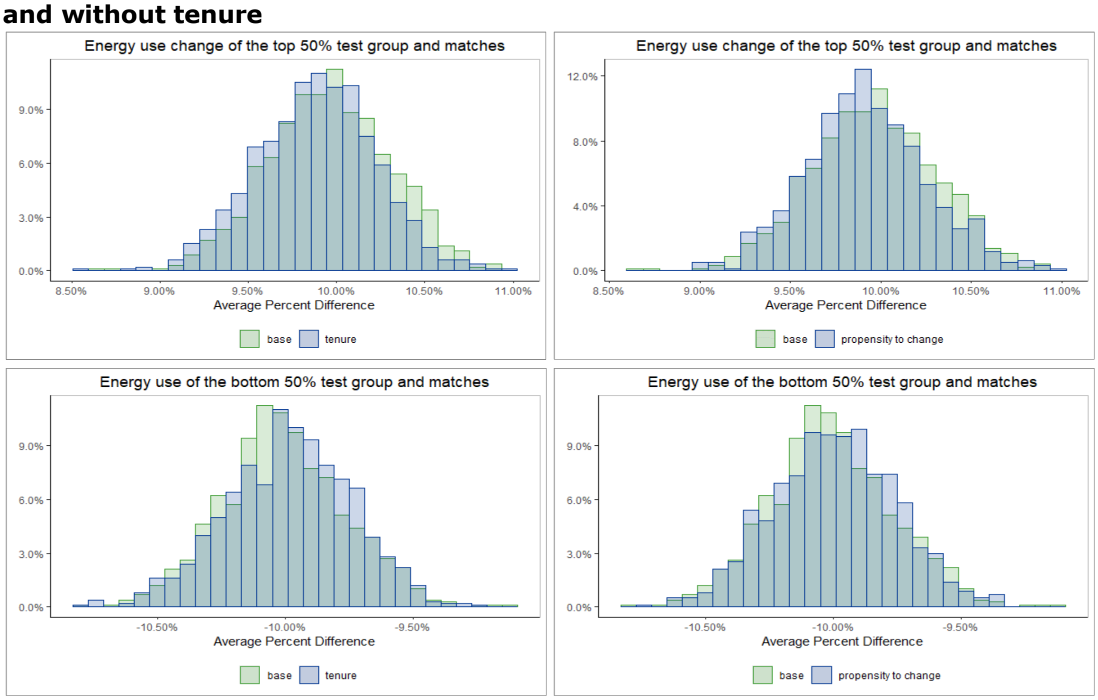
We present results from the three matching cases in Table 6-7. The table provides the mean and median post- to pre-period energy use changes for test group households with positive changes (top 50 test group) and their matches, and the same for test group households with negative changes (bottom 50 test group) and their matches. The results show marginal improvement in the mean and median changes of test groups and their matches with the inclusion of tenure; the gap between the mean change or trend of the top 50 test groups and their matches shows a minimal 2% improvement when tenure is used in matching. The simulation results suggest that tenure could provide a small improvement in bridging the gap in observed trend differentials among households.

Table 6-7. Simulation test results from matching with tenure – PG&E Electric

Matching Strategy	Metric	Top 50		Bottom 50	
		test group	matched group	test group	matched group
Base case (20% draw, 1000 simulations)	Average change	452	-196	-1009	-348
	Median change	183	-162	-737	-302
Base case + tenure (20% draw, 1000 simulations)	Average change	452	-192	-1009	-352
	Median change	183	-159	-737	-304
Base case + propensity-to-change (20% draw, 1000 simulations)	Average change	452	-203	-1007	-345
	Median change	183	-161	-736	-298

We also examined the distribution of the percent difference between the energy use changes of those matched with tenure and without tenure. The left panel of Figure 6-5 provides the distribution of percent change in energy use between those in the test groups and groups they are matched to using the base case approach and the approach that includes tenure in matching. The right panel provides the analogous for test groups and their matches matched using the base case approach and propensity to change. If the inclusion of tenure and propensity to change improve matching, the percent difference between those with extreme positive and negative trends and their matches should be lower. The distributions provided in the figure indicate that this is the case; tenure and propensity to change appear to improve matches selected for those with extreme positive and negative trends.

Figure 6-5. Distribution of the percent difference between the energy use changes matched with and without tenure



The findings presented above indicate that tenure could improve matches between participant and non-participant households by closing gaps in potential trend differentials that exist between them to a limited degree. Thus, given that tenure improves matches between households artificially set up to display extreme trend differentials, DNV GL used tenure and propensity-to-change to match participant to non-participant households to update smart thermostat savings estimates. The resulting balance in these matches are presented in Appendix D.

6.3 Appendix C: Savings Methods

Savings estimates from smart thermostat installations were based on the same two-stage modeling approach detailed in the March Report. The approach uses variable degree-day PRISM-inspired, site-level models with a matched comparison group in a difference-in-difference (DID) framework. We describe the data, matching approach, site-level weather normalization and the DID models used in this update study briefly below.

6.3.1 Data

DNV GL used data from participants that installed smart thermostats offered through PY2018 rebate programs of PG&E, SDG&E, SCE, and SCG. Table 6-8 provides the programs and number of smart thermostats rebated through them as well as the savings claimed for this measure under each program. The thermostats rebated through these programs constitute 24% of total installations with claimed electric savings and 34% of total installations with claimed gas savings.

Table 6-8. PY2018 smart thermostat rebate program by IOU

Program Name and Administrator	Program ID	Installations with electric savings claims	Installations with gas savings claims	Gross first-year savings	
				Electric (kWh)	Gas (Therms)
PG&E Residential Energy Efficiency/Plug-Load & Appliance	PGE21002	18,386	18,407	3,018,614	434,225
SCE Plug Load and Appliances Program	SCE-13-SW-001B	7,478	7,478	1,476,366	90,554
SCG Residential Energy Efficiency Program/Plug-Load & Appliance	SCG3702	9,977	41,642	1,976,966	593,205
SDG&E Plug Load And Appliances-Home Energy Efficiency Rebate	SDGE3203	2,181	1,584	337,791	15,927
SDG&E Plug Load And Appliances-Point of Sale Rebates	SDGE3204	7,363	5,442	1,149,565	54,585
Total		**45,385**	**74,553**	**7,959,303**	**1,188,496**

Three sets of participant and non-participant data were used to test the matching and/or to update estimated savings. They include:

- Non-participant and participant daily electricity and gas use data from January 2017 to December 2019 based on the initial 10:1 matches used in the March Report ("initial 10:1 matches"); these matches were obtained from propensity score matching of monthly billing data by climate zone
- Non-participant and participant daily electricity and gas use data from January 2017 to December 2019 from updated 10:1 matches conducted for follow-on work ("updated 10:1 matches"); these matches were obtained from Mahalanobis distance matching of total energy and the ratio of summer-winter energy use by climate zone and form the basis of the updated matched comparison savings estimates
- Future (PY2019) and current participant (PY2018) daily electricity and gas use data from January 2017 to December 2019 ("future participants")

Table 6-9 presents the number of PY2018 customers whose data is used to estimate smart thermostat savings. These counts are the same as those used in the March Report, with a handful less for whom we did not receive the required daily data.

Table 6-9. Customer data used in updated study

Customer Data Attrition	SCE electric	SCG gas	PG&E electric	PG&E gas	SDG&E electric	SDG&E gas
customers in 2018 tracking data with rebated smart thermostats	7,184	40,987	17,728	17,711	9,536	7,021
customers installing smart thermostats only	6,952	38,634	16,073	16,089	8,955	6,496
customers with sufficient interval data used in the analysis*	5,798	20,096	5,346	6,122	5,097	4,750

*Customers with 2018 installation date, adequate interval data for analysis (a year of pre- and post-installation), and not net metered

As indicated in the list above, the analysis is based on monthly and daily energy consumption data obtained from the IOUs. Billing data were primarily used to identify 10:1 matched comparison households while daily data were used to generate variables used in final 1:1 matching, to estimate weather normalization models, and form the basis of models used to estimate the effect of smart thermostats on energy use. Both billing and daily data were screened to remove duplicate reads, total zero energy use for the day and the year and reads that correspond to on-site solar production.

DNV GL sourced hourly weather data for 117 NOAA weather stations across California that provide historical weather observations and for which typical meteorological year (TMY) series were developed (CZ2018). CZ2018 are typical meteorological year weather data for select California weather stations that are useful for long-term weather normalization. They are provided on California's Measurement Advisory Council site and update the 2010 typical year weather data to reflect more recent weather trends.[11]

DNV GL applied the following data filtering protocols, in line with CalTRACK recommendations, and used weather data from 73 weather stations that have complete and usable data for the analysis.[12] The protocols include:

- Interpolation of gaps of up to 6 consecutive hours
- Use of daily average data only for days missing no more than 12 hourly temperature reads
- Use of data from stations that have at least 90% of the data for each year needed in the analysis

6.3.2 Matching

When treatment or intervention is applied, analysis based on an experimental design can be used to assess the impact of the intervention. In the current context, the effect of smart thermostats on energy use can be estimated using data from participants that installed the device and a comparison group. Since participants and a general comparison group are different before the intervention, it is difficult to disentangle the effect of the intervention from other differences that exists between the two groups.

Matching is a method that bridges the gap in these differences and, thereby, allow the identification of comparison group households that are like the participants along important dimensions enabling the analysis of the effect of the intervention. The matched comparison group forms the foundation of the experimental design used in this study. This quasi-experimental set up is commonly used to construct a comparison group

[11] http://calmac.org/weather.asp

[12] http://docs.caltrack.org/en/latest/methods.html#section-2-data-management

for the purposes of generating a counterfactual used to estimate the impact of smart thermostats on energy use.

DNV GL used Mahalanobis distance matching without replacement to test matching strategies and to prepare the final 1:1 matched comparison data from the initial and updated 10:1 matches used in the analysis. Future participants and one-year trend matching tests used matching with replacement because the number of non-participants available for matching was not always sufficiently large to allow matching without replacement.

Mahalanobis distance matching is scale-invariant and considers correlations of covariates to generate matches that are well-balanced. Balance is tested using standardized mean differences, the ratio of the variance of participant to matched comparison households, and visual inspection of the distribution of covariates of participants to matched comparison households. The 'base-case' variables used in the matching models included total energy use, the ratio of summer-to winter energy use, and for electricity, 6 p.m. kWh for identified 'heat wave' periods used to capture peak demand conditions. 'Heat wave' periods were identified for each climate zone as weekdays between June through September where most customers had their maximum 6 p.m. kWh.

6.3.3 Weather Normalization

DNV GL weather normalized energy consumption at the site-level based on the widely used PRISM-model. Weather normalization puts energy consumption on the same level so that changes in energy use can be attributed to factors other than weather. CalTRACK consistent PRISM-models used in the study are specified in the following way:

$$E_{im} = \beta_0 + \beta_h H_{im}(\tau_h) + \beta_c C_{im}(\tau_c) + \varepsilon_{im}$$

Where:

E_{im} - Average electric (or gas) consumption per day for participant i during period m.
$H_{im}(\tau_h)$ - Heating degree-days (HDD) that participant i faces during period m at the heating reference temperature, τ_h.
$C_{im}(\tau_c)$ - Cooling degree-days (CDD) that participant i faces during period m at the cooling reference temperature, τ_c, (not included in gas models).
$\beta_0, \beta_h, \beta_c$ – Site-level regression coefficients measuring intercept (base load), heating load, and cooling load, on a single year's energy consumption, respectively.
τ_h - Heating base temperatures, determined by choice of the optimal regression.
τ_c - Cooling base temperatures, determined by choice of the optimal regression.
ε_{im} – Regression residual.

Consumption is estimated over a range of 60°F to 80°F for cooling and 50°F to 70°F for heating to identify the temperature base or reference points for each site (household). For a base point to be considered, there needs to be at least 10 days with degree days greater than zero per year and a total of 20 or more-degree days per year, following the CalTRACK methodology.[13] Statistical tests identify the optimal set of base points. The site-level models produce parameters that indicate the level of energy consumption not correlated with either HDD or CDD (baseload), and the levels of energy consumption correlated with HDD (heating load) or CDD (cooling load). DNV GL estimated site-level models using daily data. These models were screened to remove estimates that had implausible (negative) cooling and heating coefficients.

---

[13] CalTRACK 3.2.2.2. http://docs.caltrack.org/en/latest/methods.html

Model parameter estimates for each site allow the prediction of site-level consumption under any weather condition. For evaluation purposes, all consumption is put on a typical weather basis, using CZ2018 TMY values, and produces an estimate referred to as normalized annual consumption (NAC). NAC for the pre- and post-installation periods are calculated for each site and analysis time frame by combining the estimated coefficients $\hat{\beta}_h$ and $\hat{\beta}_c$ with the annual typical meteorological year (TMY) degree days H_0 and C_0 calculated at the site-specific degree-day base(s), $\hat{\tau}_c$ and $\hat{\tau}_h$. NAC values estimated in this manner form the basis of the second stage DID models.

6.3.4 Difference-in-Difference modeling

A model based on the pre-to-post difference in NAC for participant households and a matched comparison group is estimated using a difference-in-difference modelling approach. This model is given by:

$$\Delta NAC_i = \alpha_0 + \beta T_i + \varepsilon_i$$

In this model, i subscripts a household and T is a treatment indicator that is 1 for smart thermostat households and 0 for the matched comparison homes. The effect of the program is captured by the coefficient estimate of the term associated with the treatment indicator, $\hat{\beta}$. The pre- and post-periods are demarcated by a one-month blackout period, which is based on the month of smart thermostat installation.

6.4 Appendix D: Balance in Matched Samples

We provide tests of balance for households matched using base case matching variables, base case matching variables and tenure, and base case matching and propensity-to-change in this section. We also provide tests of balance for matches between current and future participants. The primary statistics used to test balance for matched households are based on standardized means of total energy use and the ratio of the variance of the same for participant and matched comparison households.

Table 6-10 provides these statistics for the different matching strategies and samples matched based on them. The values indicate that all matched samples are well-balanced with respect to the variables used to match them across all matching strategies.

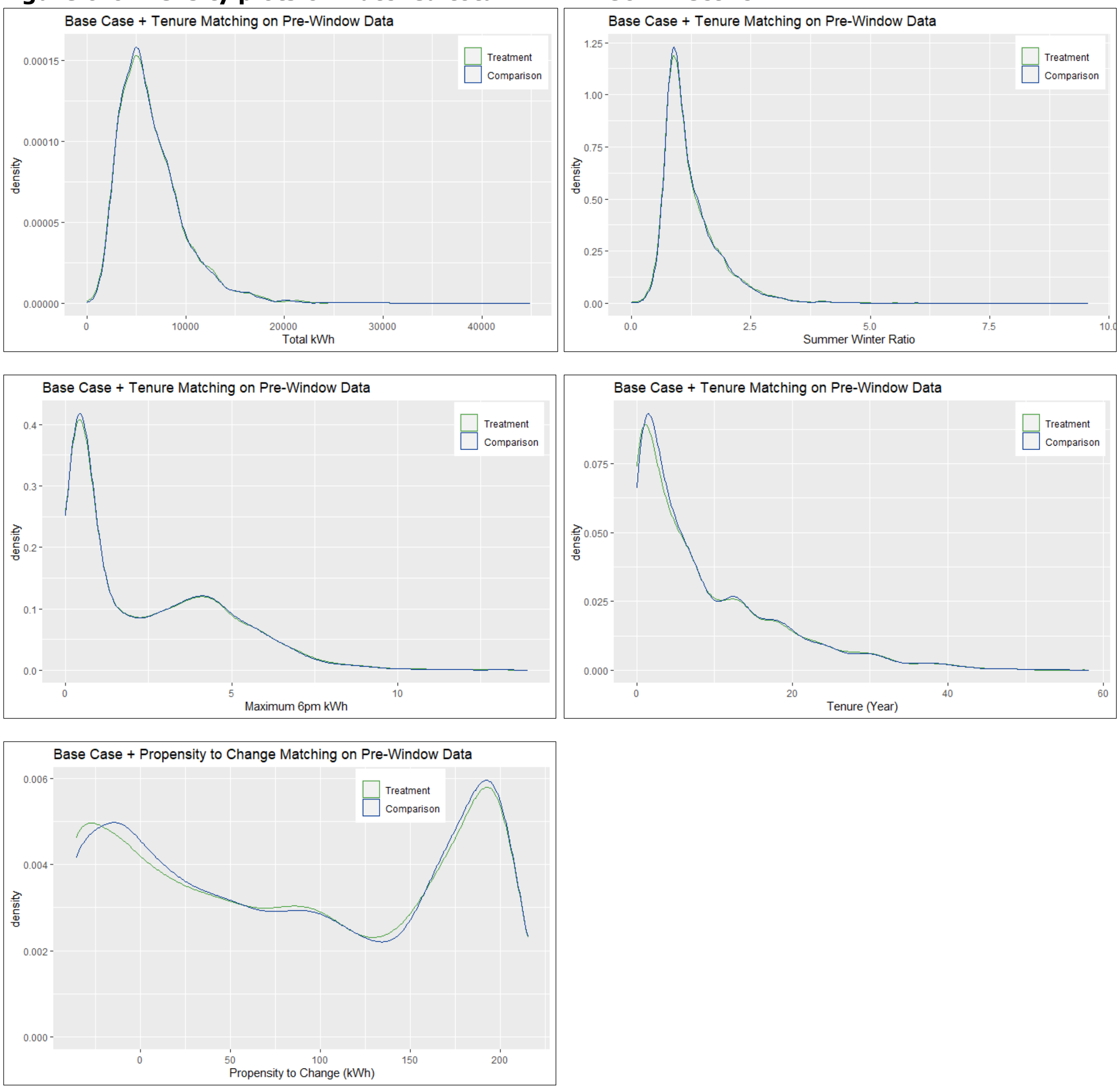
Table 6-10. Test of balance from different matching approaches

IOU	Fuel	Matching approach	Standardized difference (d)	Variance ratio (R)
PG&E	Elec	Base case model - pre-window data	0.004	1.06
		Base case model + tenure - pre-window data	0.008	1.08
		Base case model + propensity-to-change - pre-window data	0.008	1.08
	Gas	Base case model - pre-window data	0.000	1.00
		Base case model + tenure - pre-window data	0.000	1.01
		Base case model + propensity-to-change - pre-window data	0.000	1.01
SDG&E	Elec	Base case model - pre-window data	0.003	1.04
		Base case model + tenure - pre-window data	0.003	1.05
		Base case model + propensity-to-change - pre-window data	0.003	1.05
	Gas	Base case model - pre-window data	0.000	1.00
		Base case model + tenure - pre-window data	0.000	1.00
		Base case model + propensity-to-change - pre-window data	0.000	1.01
SCE	Elec	Base case model - pre-window data	0.002	1.05
		Base case model + tenure - pre-window data	0.003	1.07
		Base case model + propensity-to-change - pre-window data	0.002	1.07
SCG	Gas	Base case model - pre-window data	0.000	1.00
		Base case model + tenure - pre-window data	0.000	1.01

IOU	Fuel	Matching approach	Standardized difference (d)	Variance ratio (R)
		Base case model + propensity-to-change - pre-window data	0.000	1.01

DNV GL also investigated balance visually using density plots of variables used in matching. Figure 6-6 provides an example of density plots of matched total energy from different matching strategies using data from PG&E electric customers. Density plots for the remaining matching variables, other IOUs and fuel provide similar results.

Figure 6-6. Density plots of matched total kWh - PG&E Electric



All the figures indicate that each matching strategy provides balanced samples, but they don't make it possible to identify if any one matching strategy is superior in terms of bridging the possible trend differential that exists among participant and non-participant homes. Results from approaches to gain some insight into this issue are summarized in section 4 of the report.

6.5 Appendix E: Response to comments

Comment #	Commenter	Page (in the document); or "Overarching" for general comments	Comment/feedback/change requested	Evaluator's response
1	SDG&E	4	COVID can also be pointed to as a future motivation for increased energy use after a smart thermostat has been installed.	Yes, though an ideal comparison group would also pick up the increase and control for it. Self-selection could still be an issue as participants and nonparticipants respond differently to COVID. To be clear, though, the current evaluation is based on installations and data prior to COVID. Thus, SCT related energy consumption changes in the current evaluation are unaffected by it.
2	SDG&E	6	When comparing 2018 vs 2019 smart thermostat rebates, it should be noted that qualification changes opened up a larger selection of smart thermostats - some cheaper and more economical. This may have changed the type of person who is purchasing a thermostat - moving away from young and affluent towards a more diverse group of people.	Noted. DNV GL will be evaluating SCTs delivered via direct install programs but not those installed via rebates for PY2019. We will examine if the type of customers that received SCTs via direct install programs are similar to those described in the comment and what effect that has on energy consumption changes. Further, although the PY2019 SCT impact evaluation will not be based on rebated smart thermostats, DNV GL will include PY2019 rebate customers in its survey and will seek to establish the demographic characteristics of these customers.
3	SCE	Overarching	Do you have any thoughts and/or reasoning as to why there is an increase in baseload on the measure case? Can this be attributed to fan-runtime as enabled by the smart thermostat?	We had indicated this to be one of the possible explanations for the measured baseload increase. In our March Report, we noted that "Major smart thermostat models offer the option of setting a daily timer on the system ventilation fan while setting up other system default settings. As a new functionality not available on most programmable thermostats, use of this capability would likely increase consumption generally, and an increase due to a regularly scheduled fan would show up in the baseload portion of the estimate." Despite this, the adjustment that was applied assumes that all of this increase is related to self-selection.
4	SCE	Overarching	Do you have any thoughts on the study's hypothesis of differential energy consumption trends between participants and their matched comparison group?	The differential energy consumption trends between participants and their matched comparison group could be tied to higher prevalence of energy use increasing actions among the former such as changes in dwelling size, household occupancy, but also other unobserved characteristics that drive households to choose to participate in a program offering smart thermostats. There is a higher prevalence of households with such characteristics among the participant group as indicated by survey results.
5	SCE	Overarching	Is there enough data to (with a high degree of certainty) conclude that participants undertook more load building actions than non-participants they were matched to; for	DNV GL noted statistically significant differences in net energy use increasing actions among participants compared to matched comparison group survey respondents in the March Report. This finding is robust as it is based on data

Comment #	Commenter	Page (in the document); or "Overarching" for general comments	Comment/feedback/change requested	Evaluator's response
			instance, they were more likely to purchase an EV than non-participants?	from about 2,400 matched comparison group households and about 10,100 participants. Specifically, program participants reported adding EV charging to the home in significantly greater proportions than non-participants at 7% to 4% respectively. The impact evaluation developed savings estimates for rebate program participants in PY 2018, and this group of customers reported adding EV charging at 8% compared to 4% for non-participants. Table 4-8 in the March Report provides further details on changes in homes that impact energy use and the differences in these between participants and non-participants.
6	Google	Self-selection	A key statement on page 4 of the report summarizes the primary remaining issue with self-selection bias: "In fact, the only limitation of a comparison group composed of future participants, with respect to controlling for self-selection, is if the increase in consumption or trend is directly or indirectly correlated with the specific installation timing of the thermostat itself. Practically, this would mean that participants are more likely to purchase an EV within 12 months on either side of the thermostat installation date than at other times. While this could still be the basis for self-selection bias in the estimates, it is part of a much narrower range of possibilities." Although it's certainly true that self-selection bias should be reduced by using future participants, the timing issue seems quite likely to still be a problem. For example, is it reasonable to assume that customers who got a smart thermostat rebate in 2018 are just as likely to purchase other smart home tech (e.g., smart speakers, doorbells, webcams, mesh WiFi) in 2018 as customers who got their rebates a year later? Households are on an evolving journey of adding tech, buying EVs, making home improvements, and having babies. How likely is it that the decision to buy a smart thermostat has no relationship in time to any of these other decisions? The main evidence provided against the commonsense expectation that timing should matter is that the prior adjustment approach and this new comparison group approach produce generally similar savings and that the future comparison group approach reportedly doesn't show any net baseload usage increase (which was taken as evidence of bias in the March report). But the new report provides no details about these changes and it therefore remains difficult to suss out potential bias. The future-participant comparison group approach was used in a recent evaluation in Illinois and that study still found evidence of increased baseload -- made especially clear by comparing load shapes on mild days. In a prior study in IL that used a matched comparison group, they reported an	Google agrees that a future participant comparison group is an improvement over a non-participant comparison group but states that self-selection "seems quite likely to still be a problem". In the evaluation, we state that the future participant comparison group decreases or removes the likelihood of self-selection in the estimates. Alternatively, if there was self-selection in the March results, the future participants would decrease or even zero out the magnitude of that self-selection bias. The fact that the self-selected population is "on an evolving journey of adding tech, buying EVs, making home improvements, and having babies", as Google says, indicates to us that their general upward trend in consumption lasts for many years, likely on either side of their thermostat purchase. This would support our argument that the future participants address the self-selection issue. The timing argument only supports the presence of self-selection (trend differential) if one can claim that there is an unusually large increase in consumption that occurs at or near the time of the thermostat installation, and not a year earlier or later. We continue to be interested in receiving evidence that EV purchase or babies have a greater concentration near the time of thermostat installation.

Comment #	Commenter	Page (in the document); or "Overarching" for general comments	Comment/feedback/change requested	Evaluator's response
			average baseload increase of 114 kWh/yr. (nearly identical to the 112 kWh/yr. in the March CA report). The newer IL study based on future participants found the baseload increase declined to about half to 37-88 kWh/yr. (depending on analysis method). The fact that the new CA analysis reports no baseload usage increase may be due to inherent uncertainties or limitations in the billing analysis approach rather than truly showing no change.	
7	Google	Self-selection	But, in any case, although an increase in baseload use may be considered evidence of self-selection bias, the lack of such a finding does not mean there is no bias. There may be an alternate narrative that can explain the findings in both reports and still have self-selection bias. One alternative explanation is that the baseload usage increase from self-selection is being partly or fully offset by a decrease in baseload usage caused by the thermostat. The new report continues the speculation from the March report that the HVAC fan duty cycle feature in smart thermostats may be responsible for the observed increase in baseload electric use. In our comments on the March report, we noted: Nest's fan scheduling feature was designed to help customers save energy by reducing the duty cycle of the fan for customers who want to run their HVAC fan in recirculation mode. A significant minority of households use their HVAC fan for continuous recirculation for all or part of the year -- typically to reduce temperature imbalances or filter the air more frequently (see studies in MN and WI). DOE estimates that HVAC fans run an extra 400 hours per year on average due to this behavior. Data from all Nest thermostats installed in California in 2018 found an average of just 105 hours of fan-only runtime in 2019 much less than the DOE estimate. This difference can be expected to save an average of 100-200 kWh/yr. (more if duct losses are included). These savings would appear in the billing analysis as a combination of baseload and cooling savings because fan-only hours are somewhat seasonal. We can provide more details if needed. We would like to reiterate this comment and provide some more evidence. The DOE document was based on data primarily from studies in Wisconsin and Minnesota. We calculated our customer fan-only runtimes in each of those states and compared them to the underlying studies used by DOE. We found that Nest customers use their fan 38% fewer hours than the WI study found and 56% fewer hours than the MN study found. It appears that the goal of the fan duty cycle control -- to give customers more efficient options for how to control their fan -- has been successful.	Google states that "There may be an alternate narrative that can explain the findings in both reports and still have self-selection bias" and they re-open the discussion related to fan-usage and implications for savings, self-selection, etc. We are not sure we follow the logic of the specific argument regarding fan usage. (The scheduled fan usage that we point to would show up in baseload, precisely because it is regularly scheduled. We do not understand why the fan usage that google claims to be saving by more efficiently using the fan around HVAC system usage would show up in baseload, as claimed.) We are interested in Google's statement, in the discussion of timing, that thermostats lead to the purchase of "other smart home tech (e.g., smart speakers, doorbells, webcams, mesh WiFi)" and their associated energy usage. That sounds like a narrative that would provide a thermostat-related explanation for why savings would be, and should be, lower than expected and not support for the presence of self-election.

Comment #	Commenter	Page (in the document); or "Overarching" for general comments	Comment/feedback/change requested	Evaluator's response
			If HVAC fans in CA are running 100 hours/yr. less due to the fan duty cycle control, then that would increase usage by about 40-50 kWh/yr. which would mostly appear as baseload use. These savings would have resulted in the March study under-adjusting for bias and would also result in the current study not finding any baseload usage increase. This alternative explanation is further supported by the customer survey results in the March report. The survey found surprisingly large differences in many exogenous energy related changes in participants' homes such as more EV charging stations and increases in household size. Reasonable estimates for the impact of these differences suggest that the 112 kWh/yr. baseload usage increase reported is smaller than expected. Baseload savings from the fan control would have offset some of this increase, explaining why that baseload change appears too small. We believe that this fan-related explanation is not only reasonable but is more likely than the alternative of assuming no temporal aspect to purchase decisions. Of course, even without the fan explanation, there still remains potential bias such as described in the report -- such as households having a baby and then adjusting thermostat settings.	
8	Google	Interpretation and application of the results comments	The realization rates reported in the study are specific to the rebate programs as operated in 2018 and to the level of savings claimed by the PAs at that time. The reported results would not apply directly to programs using different designs or where savings were claimed based on the current workpaper savings estimates. We have identified three key issues that would need to be addressed: 1. Claimed Savings: Based on data in the report, PA's claimed an average savings of 218 kWh per thermostat. But, according to calculations based on data in the report, using the older workpaper values (SCE17HC054.0) should have resulted in 190 kWh of claimed savings, indicating that part of the low realization rate was apparently savings claims that were too large. The savings based on the updated workpaper (Rev 1) would be about 162 kWh, increasing the realization rate from 33% to 45%. 2. Installation Rate: The reported savings include customers who received a rebate but did not install the thermostat in the home of record. These impacts are appropriate to include in the evaluation of the rebate impacts but should not be applied to other program designs, such as direct install, where the thermostat installation can be confirmed. Nest recently checked activation rates for thermostats that received rebates in Illinois and found that 18% of rebated thermostats did not appear to be activated in the service	Google offers three comments regarding the interpretation and application of the results. 1. Regarding the number used as a denominator for realization rate - we used the claimed savings from the tracking data, and that is the appropriate number for the denominator. We are reporting the realization rate for these programs relative to their claimed savings. 2. As noted by Google, installation rate is not a relevant consideration for the realization rate that we report in this evaluation. The savings from this evaluation were not recommended for application to units installed via the direct install channel in the subsequent workpaper for exactly this reason. 3. As noted by Google, presence of cooling is also not relevant to the realization rate that we report in this evaluation. We recognize that these considerations will produce different estimates of realization rates that would be appropriate for different populations. This was outside the scope of this evaluation. Furthermore, while illustrative, values from Chicago and/or Nest's full population of thermostats in CA are not directly relevant to this evaluation population so we do not necessarily agree with the conclusions put forward.

Comment #	Commenter	Page (in the document); or "Overarching" for general comments	Comment/feedback/change requested	Evaluator's response
			territory prior to the start of the post analysis period (7% activated outside of IL, 3% never activated, 8% activated during or after post period, see report footnote 1). If these values applied to California, then the savings would be 1.22x larger per actually installed thermostat. 3. Presence of Cooling: The reported savings include customers that don't have central air conditioning or don't use any cooling. These effects are appropriate to include in evaluating the rebate program but should not be applied to program designs that can verify the installation of the thermostats in appropriate circumstances. In the main report, 28% of the participants had their savings set to zero because they had no detectable cooling load in their billing data. Nest data from thermostat activations in California in 2018 found that 9% of the thermostats are not connected to a central air conditioner. So, the estimated savings in homes that have central cooling should be estimated at 1.1x the value reported and the savings for homes that actually use cooling would be 1.39x the savings reported. The three issues described above result in an estimated realization rate that is nearly double that reported in the study (60% vs 33%) for a program that can confirm installation of the thermostat in homes with central AC and that uses Rev 1 of the workpaper.	
9	Google	All thermostats are now optimized	Lastly, the smart thermostat market in California has changed significantly since 2018, in that two of the largest manufacturers, Google Nest and ecobee, now offer thermostat optimization (TO) programs free to all customers. (Google Nest offers Seasonal Savings to its customers while ecobee offers eco+ to its customers.) In 2020, Google Nest deployed their Seasonal Savings program to over 300,000 devices in California. These TO programs provide incremental savings that are not included in this evaluation, but which should be in any future smart thermostat saving claims in California.	DNV GL is happy to hear that optimization is now standard and free to all customers. This will address the very real possibility of double counting of savings when thermostats and optimization are considered separately. Please provide us with any new studies that provide savings estimates for new, optimized savings.
9	ecobee	Addressing self-selection bias	ecobee finds the following assertion in the report to be speculative: For example, increases in baseload consumption among participants could reflect increase in usage of ventilation fans facilitated by the scheduling features of smart thermostats. ecobee recommends for DNV GL to work with smart thermostat vendors to better understand the impact of fan settings on baseload usage by reviewing aggregate vendor telemetry data specific to the devices used in the analysis. ecobee believes it is more likely that the increase in baseload consumption observed in this analysis is a sign of	Regarding fan usage discussion being "speculative": The discussion of fan usage in the DNV GL evaluation report is intended to be speculative with the intent to show that there are scenarios that could explain the results. The statements that the results are affected by self-selection bias are similarly speculative. The method used remains the industry standard approach to estimating savings and the inclusion of a future participant comparison group further established these as the best possible estimates of savings. The statement that telemetry data should be used indicates an ongoing misunderstanding regarding the appropriate baseline for this analysis. The existing thermostat, likely programmable or even manual, is the baseline condition and would not have telemetry data for the purpose of an

Comment #	Commenter	Page (in the document); or "Overarching" for general comments	Comment/feedback/change requested	Evaluator's response
			self-selection bias and should not be attributed to smart thermostats. Due to such self-selection bias issues inherent in quasi-experimental design, ecobee does not believe that quasi-experimental methods are a suitable alternative on their own to randomized experiments for smart thermostat evaluation, and that the use of quasi-experimental methods for smart thermostat evaluation should be strongly discouraged for policymaking purposes.	evaluation. Regarding "self-selection bias issues inherent in quasi-experimental design": This is a false statement. The quasi-experimental design used provides the best feasible set of controls for self-selection. The approach cannot claim to necessarily avoid all self-selection, but this is fundamentally different than claiming it is an inherent issue. Support for claims of self-selection need to be based in the data. Other than those which we discuss in our report, we have seen no further data supporting a claim of self-selection. The interim results, for which these comments apply, use a future participant comparison group that more completely addresses concerns regarding self-selection and produces similar results to the March findings.
10	ecobee	Addressing self-selection bias	In this case where a quasi-experimental design has already been utilized, ecobee believes that the best approach to address self-selection bias more directly and effectively is to run the ENERGY STAR metric on the devices used in the study and adjust for setback behavior. ENERGY STAR uses telemetry data and measures only HVAC changes, while whole-house data measures all changes made to HVAC and non-HVAC related end-uses alike. Therefore, the ENERGY STAR metric would not incorrectly attribute a baseload increase to smart thermostats. By comparing and averaging the two methods, DNV GL could better assess the true savings value. This approach has been recommended by the Northeast Energy Efficiency Partnerships (NEEP).	Regarding use of the ENERGY STAR metric: We have stated numerous times that the ES metric was not created to be used in evaluations and is not appropriate for evaluation purposes. It uses a constructed, unrealistic baseline that ignores the existing thermostat.
11	ecobee	Including optimization savings	ecobee recently launched a new thermostat optimization platform called eco+ that has been released to all of its thermostats in the form of a free software upgrade. Leading up to the release, ecobee contracted Demand Side Analytics, a third-party measurement and verification firm, to measure the climate-specific additional energy and demand savings impacts of this software upgrade through a robust randomized encouragement design using nearly 250,000 ecobee devices. The executive summary of the full measurement and verification report is attached. As being done in other regions, the additional optimization savings presented for the California-specific climate zones, Dry and Marine, should be added as of summer 2020 onwards.	Ecobee notes that optimization has become standard for all ecobee thermostats and the savings should be added going forward. We will note this is future ex ante savings considerations. The DSA report, as provided, has insufficient detail to fully understand the results of the study. Furthermore, these savings are not necessarily additive to thermostat replacement savings. Finally, the energy efficiency savings in the attached report are limited and appear to be mixed with TOU effects. Ecobee should provide evaluation evidence that the new optimized ecobee thermostats, when replacing an existing thermostat, produce savings.