

CalTRACK and Commercial Buildings: Confronting COVID and Other Non-Routine Events

*Stephanie Bieler, Josh Bode, and Savannah Horner, Demand Side Analytics
Brian Smith and Lauren Garcia, Pacific Gas and Electric*

ABSTRACT

Population-based NMEC—a savings estimation methodology that’s appealing for its economy to implement and its ease in accommodating measures of all types—is being adopted increasingly by program administrators. One utility that selected the Population NMEC measurement methodology—short for normalized metered energy consumption—enrolled over one hundred commercial and governmental customers into an on-bill financing program in 2019. When shelter-in-place orders were issued early in the monitoring period, a missing element of CalTRACK—the open-source population NMEC code—became apparent to all: there wasn’t a method developed to incorporate comparison groups into the calculations as a means to control for the systemic “non-routine event” brought by the pandemic in how, when, and how much electricity is used. Settling pay-for-performance contracts and claiming savings accurately were threatened, and data privacy restrictions prohibited the utility from providing non-participant customer data to third parties to construct comparison groups readily. We assessed the performance of regression models, matching methods, and other statistical techniques for constructing accurate and unbiased comparison groups without revealing personally-identifiable information. The study investigates:

- How COVID affected non-residential loads
- How pre-post NMEC methods without a comparison group performed under pandemic conditions
- How comparison group methods that used aggregated data compared to those that used customer-level data
- Whether comparison groups with non-participant data are necessary or if groups that use aggregated profiles perform well

Introduction

The COVID-19 pandemic led to changes in our commutes, business operations, and home use patterns. It changed how, when, and how much electricity households and businesses consumed. Moreover, the impact of COVID-19 on non-residential electricity use fluctuated over time as different sectors of the economy alternately shut down, reopened, and experienced shifts in activity levels due to the multiple waves of COVID variants, changes in social distancing practices, and vaccine availability. While the effects differed by sector, fuel type, and geography, COVID typically increased residential energy consumption and decreased commercial and industrial consumption compared to pre-COVID periods.

Figure 1 shows the estimated percent change in non-residential electricity consumption by date as a result of the pandemic for key non-residential industries in Pacific Gas and Electric (PG&E) territory. The estimates of change in energy use were developed by using a random sample of 50,000 PG&E customers, constructing weather-normalized models of energy use from

March of 2018 through February of 2020, and then predicting what consumption would have been in each segment for the post-COVID period starting in March 2020.

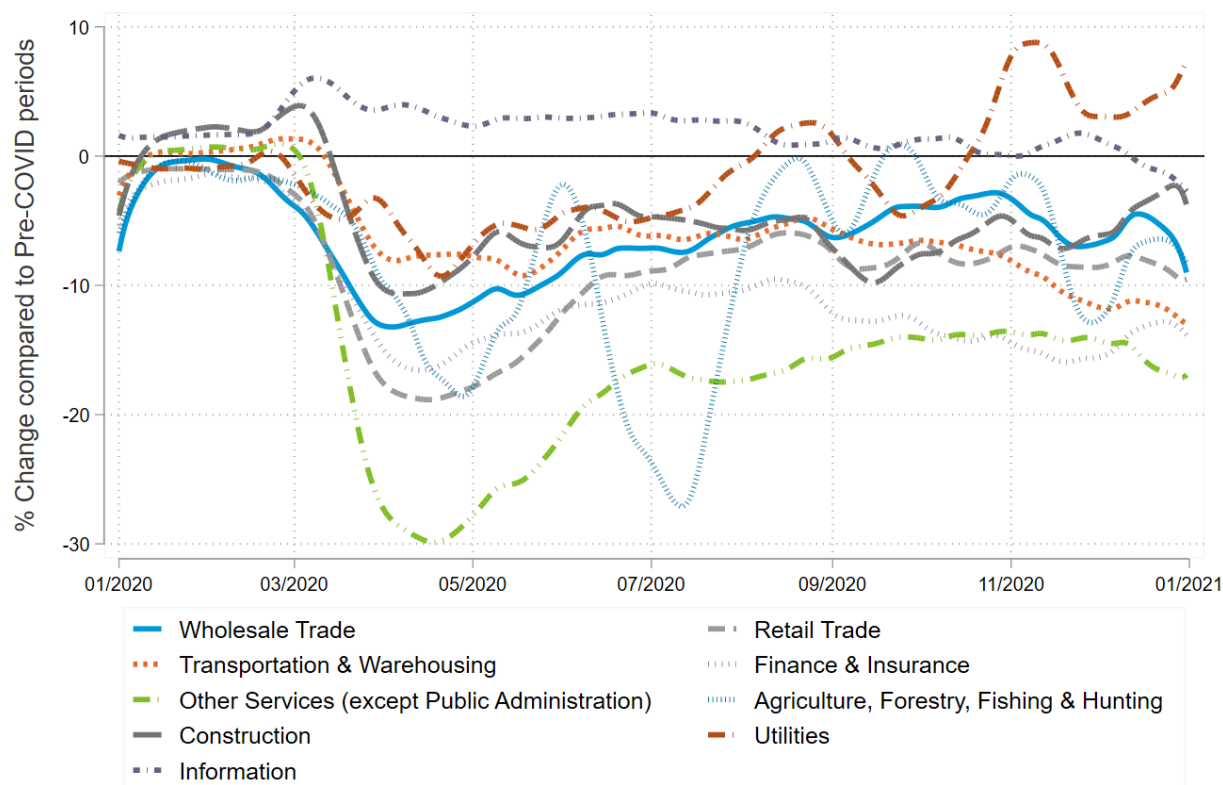


Figure 1: Commercial Changes in Energy Consumption Associated with the Pandemic

As expected, Retail, Offices (Finance & Insurance), and Other Services showed the highest monthly decreases in consumption (Table 1). On average, non-residential customers saw approximately an 8-10% decline in consumption associated with the pandemic. Compared to the residential sector, the changes in consumption were slightly more variable, due to local and statewide changes in guidance about which business types were allowed to operate and in what capacity.

Table 1: Monthly Commercial Changes in Energy Consumption Associated with the Pandemic

Year Month	Agriculture, Forestry, Fishing &	Utilities	Construction	Wholesale Trade	Retail Trade	Transportation & Warehousing	Information	Finance & Insurance	Other
2020-01	-1.4%	-0.8%	0.9%	-1.3%	-1.1%	-0.2%	1.5%	-2.2%	-0.1%
2020-02	-1.2%	-0.3%	2.0%	-1.1%	-1.1%	0.7%	1.8%	-1.6%	0.6%
2020-03	-4.6%	-4.3%	-1.6%	-9.0%	-10.0%	-2.1%	5.4%	-9.0%	-12.5%
2020-04	-16.3%	-7.4%	-10.4%	-12.6%	-18.7%	-7.9%	3.1%	-16.5%	-29.6%
2020-05	-10.7%	-5.2%	-6.7%	-10.2%	-15.7%	-8.5%	3.1%	-13.0%	-25.1%
2020-06	-12.1%	-4.5%	-4.1%	-7.5%	-9.9%	-5.5%	3.1%	-11.1%	-17.8%
2020-07	-22.2%	-3.4%	-5.7%	-6.9%	-7.8%	-6.4%	2.9%	-10.7%	-17.1%
2020-08	-1.9%	1.7%	-4.2%	-4.8%	-6.3%	-5.3%	1.4%	-9.7%	-16.2%
2020-09	-2.7%	-2.2%	-9.8%	-5.4%	-8.3%	-6.7%	0.7%	-12.8%	-14.9%
2020-10	-3.5%	-0.4%	-5.6%	-3.0%	-7.6%	-6.9%	1.0%	-13.9%	-13.6%
2020-11	-7.3%	7.2%	-6.9%	-6.2%	-8.1%	-10.2%	1.0%	-15.5%	-14.4%
2020-12	-7.0%	4.1%	-3.7%	-5.2%	-8.2%	-11.4%	-0.4%	-13.6%	-15.3%

The COVID-19 pandemic laid bare the limitations of pre-post methods that attempt to estimate energy savings without a control group. Methodologies that do not include a comparison group are more prone to confound non-routine events – such as a pandemic – with the impact of the intervention – energy efficiency. Moreover, the effect of the pandemic on energy use will likely persist because it has fundamentally altered commute and work-from-home patterns. We need to account for more than just the timing of the installations and weather to measure the energy efficiency impact accurately. The pandemic has highlighted the need to incorporate comparison groups to calibrate estimates of energy savings for programs that typically rely on a simple pre-post methodology.

Overview of PG&E’s On-Bill Financing Program

PG&E’s On-Bill Financing-Alternative Pathway (OBF-AP) program is a non-residential program that offers customers 0% interest loans to finance the purchase and installation of energy efficiency (EE) measures. It’s referred to as an “alternative pathway” because customers can access on-bill financing without the usual requirement of participating in rebate and incentive programs. OBF-AP participant customers repay the loan principal as a fixed monthly charge on their PG&E bill. The OBF-AP loan is sized based on project cost and considers the projected energy savings at the customer’s facility. Commercial customers can receive a loan of up to \$250,000 based on ten years of projected energy savings; government agency customers can receive up to \$250,000 for ten years of projected energy savings (or up to \$4,000,000 for unique energy savings opportunities). The program is designed to remove the barrier of up-front costs and financing for energy efficiency measures. The loan is required to be repaid in full by the participating customer, with the threat of utility disconnection for non-payment.

OBF-AP measures primarily consist of lighting, refrigeration, and controls. A total of 139 customers received EE interventions in 2019, generating first-year savings through 2019 and 2020. The majority of interventions took place in the second half of the 2019 calendar year with first-year savings occurring in 2020. The zero interest loans offered by PG&E are based on ex-ante estimates informed by engineering calculations, which provide an estimate of the magnitude of energy reductions we can expect from each customer.

Table 2 summarizes the expected savings by project type divided into two broad categories, lighting and non-lighting. We also include the annual usage (before energy efficiency) to understand the expected percent change in energy use. We can see that the majority of the customers and expected savings come from lighting projects. On aggregate, the expected percent reduction from these projects is 22%. While the ex-ante expected savings are large, the period when first-year energy savings are calculated overlapped with a large non-routine event – the COVID-19 pandemic.

Table 2: Expected Electric Energy Savings of 2019 Participants

Project Type	Participants	Participant Pre-Treatment Annual Usage (GWh)	Expected Savings (GWh)	Expected Savings as % of Historic Usage
Lighting	117	31.1	8.7	28.0%
Non-Lighting	17	21.3	2.6	12.3%
Both	5	6.4	1.5	23.0%
Total	139	58.8	12.8	21.8%

Figure 2 examines the distribution of customer size and expected savings for the portfolio of 2019 OBF-AP participants. In each of the stacked bars, one bar segment represents 20% of the 2019 OBF participants (or one quintile). The first bar represents the percent of customers in the OBF-AP population, so each bar segment or quintile is the same size as each represents 20% of the OBF-AP customers. The second bar depicts each quintile's contribution to the OBF-AP population's total pre-treatment consumption. The third bar depicts each quintile's expected contribution to the PY2020 savings. The expected savings and electricity consumption (before energy efficiency) are highly concentrated among the largest OBF-AP participants. The top 20% of customers account for almost 70% of the population's electricity use before energy efficiency and 70% of the expected energy savings.

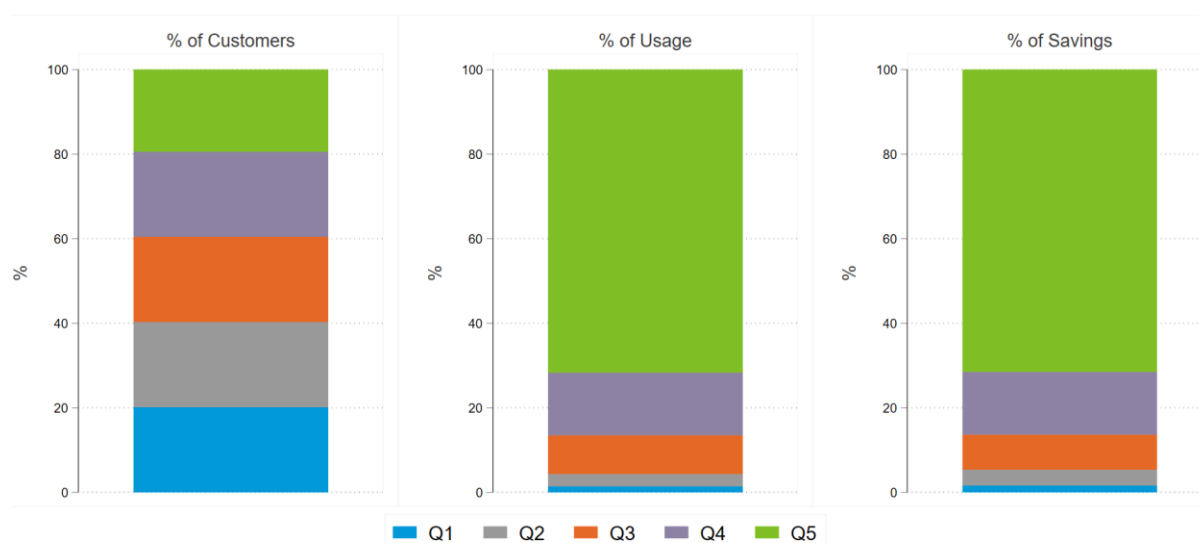


Figure 2: Participant Size and Expected Savings Distribution of 2019 Participants

NMEC Methodology Overview

Historically, and at the simplest level, producing the energy savings using NMEC has involved three steps:

1. Use the data from the pre-intervention period to develop a model of energy use as a function of temperature, hour of week, and estimated occupancy.
2. Apply the model to the post-intervention period to produce a baseline.
3. Estimate the savings as the difference between the baseline and actual energy use during the period after the energy efficiency intervention took place.

However, COVID-19 highlighted the limitations of pre-post methods without a control group. To account for the exogenous impact of COVID, we held a tournament to test the accuracy of multiple methods of incorporating comparison groups. Figure 3 outlines the process used to identify the most accurate approach to estimate savings.

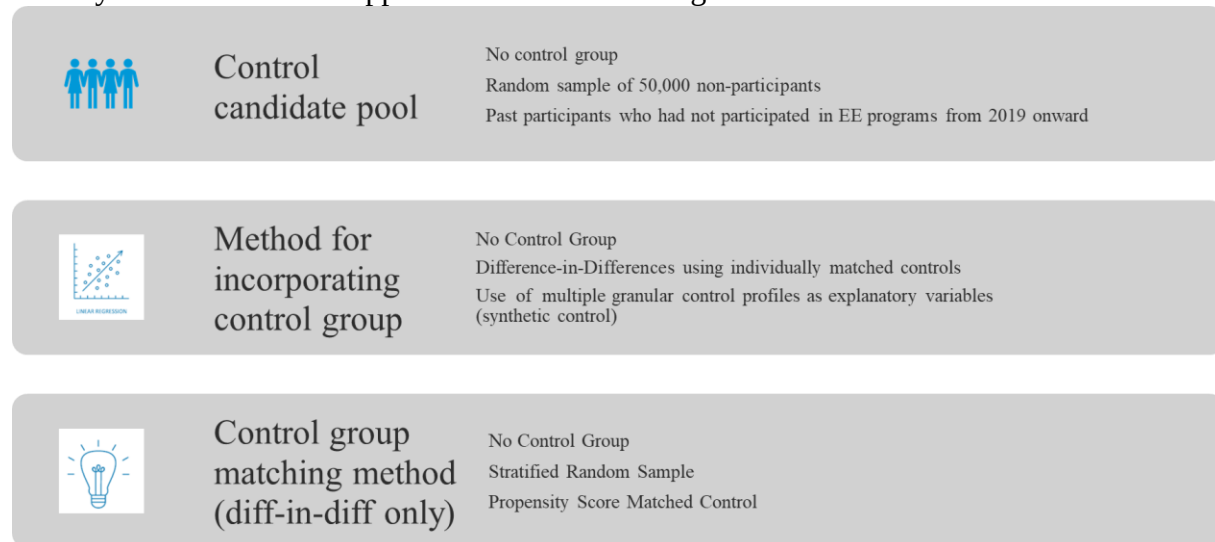


Figure 4 describes the methods tested in accuracy tournament. In addition to existing pre-post method without a comparison group, the tournament assessed how accuracy varies based on the pool of control candidates used, how the matched control group was selected, and the method used to incorporate the control group.

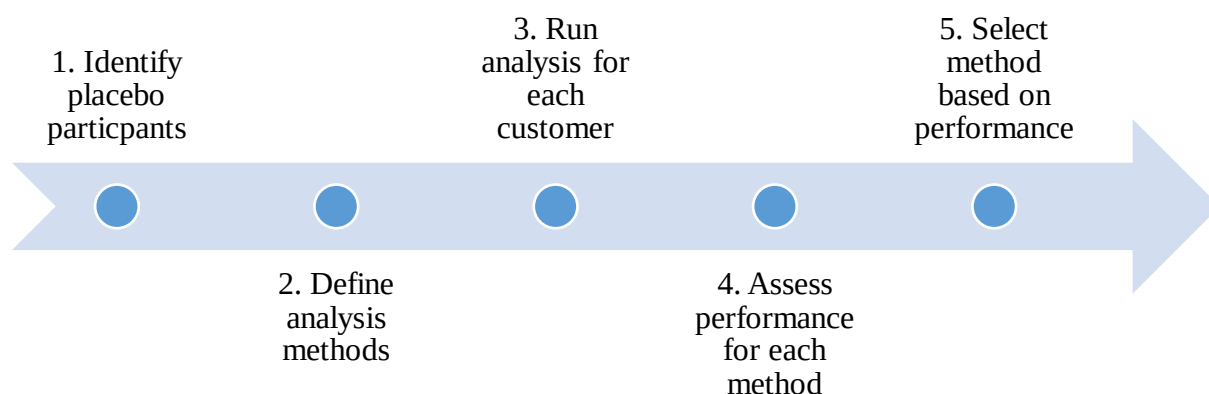


Figure 3: OBF-AP Evaluation Framework

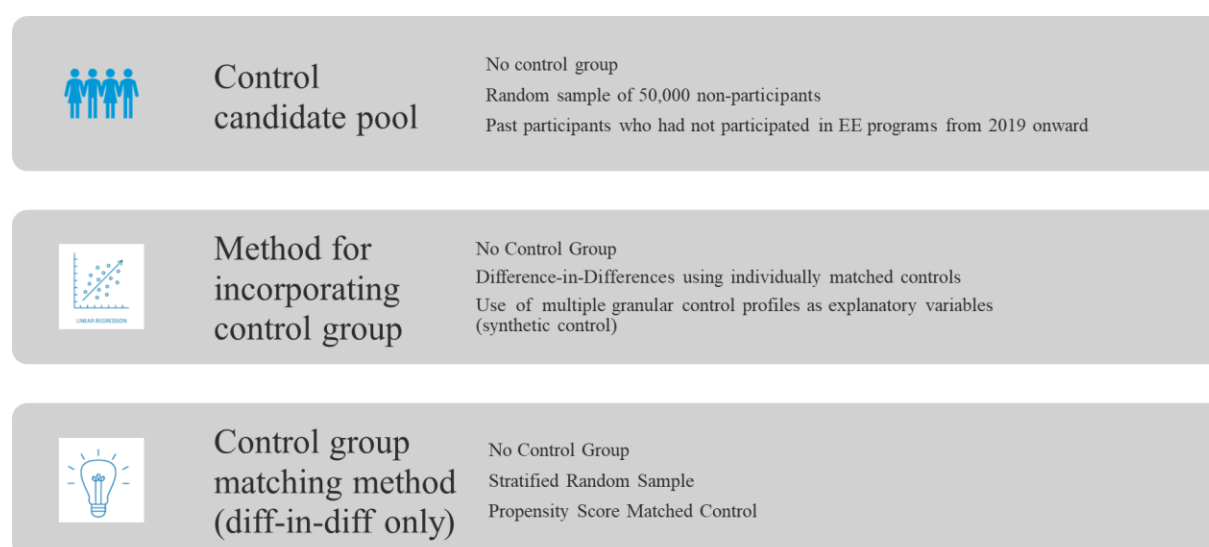


Figure 4: Control Methods Tested in Accuracy Tournament

We used three control candidate pools:

1. **No comparison groups.** This simple pre-post method is used as a baseline to test for how the model performed without any controls. It reflects the how energy savings have historically been estimated using NMEC methods.
2. **A random sample of non-residential customers.** The first pool of candidates consisted of a random sample of non-residential customers that had not participated in EE programs after 2015 from the PG&E territory. The total sample pool consisted of 50,000 customers that were stratified based on the prior year's electric consumption (kWh).
3. **Past Participants.** The second pool of candidates consisted of a random sample of non-residential customers that had participated in EE programs in the 2017-2018 period but did not participate in programs from 2019 onwards. These customers are most like actual participants in EE programs because they were past participants in EE programs.

Selecting a comparison group composed from this group helps control for self-selection bias: the non-visible characteristics that drew them to participate in EE programs in the recent past may be similar to those that drew customers to participate in the 2019 OBF-AP program.

We used three approaches for incorporating comparison groups into our analysis:

1. **No comparison groups.** The pandemic's impact on energy usage underscores the necessity of incorporating a comparison group. This simple pre-post method is used as a baseline to test for how the model performed without any controls.
2. **Using a difference-in-differences approach with individual matched controls.** In this approach:
 - a. The regression method is run for all participants and their matched comparison groups.
 - b. The difference between the estimated baseline and actual usage for the control group is error since they did not participate in the energy efficiency program.
 - c. The errors observed in the comparison group are then subtracted from the impacts estimated for the participant group.
 - d. This difference-in-differences calculation produces an adjusted impact, net of non-routine events and other variables not captured by the regression models.
3. **Adding granular control profiles to the method as a right-hand-side/explanatory variable.** We refer to this method as the "Aggregated Control" method.
 - a. This approach relies on an aggregated profile of non-participant usage in the pre- and post-period to explain additional variation in customer consumption patterns. Each profile is effectively an index of energy usage patterns.
 - b. We developed granular profiles from the 50,000 PG&E non-participant customers. To develop the aggregate profiles, customers were segmented based on their climate zone, solar status, rate class, and industry. We then aggregated the segments to calculate the average usage for each segment across the entire analysis period. Each customer was assigned up to 14 industry profiles based on their climate zone, rate class, and solar status.
 - c. The aggregated profiles are added as explanatory variables (right hand side variables) in the existing regression method. The regression effectively weights the aggregated non-participant profiles based on predictive power to produce a counterfactual.
 - d. This method protects the confidentiality of non-participant data as it is aggregated and can be provided to third parties and energy efficiency aggregators for validation.

For the difference-in-difference approach, we developed a matched comparison group for the treatment customers. Each treatment customer was matched to a single control customer. For a comparison group to be a good proxy, evaluators should ensure that the controls are statistically similar to the participants on observable characteristics. We tested two methods for selecting a matched control group:

1. **Stratified random sample:** For each participant, randomly select a non-participant to act as a control within a grouping of segmentation characteristics. For example, non-participants must be matched to participants in the same climate zone and industry classification.
2. **Propensity score matching:** Within each segment of interest, fit a probit model to estimate the probability of enrollment as a function of observable characteristics and pre-treatment usage patterns. The model is used to produce the adoption propensity, given the characteristics included in the model, and non-participants are matched to participants based on their propensities to participate in the EE program. The approach is effectively identifying a statistical look-alike based on observable variables.

Assessing Accuracy

To assess accuracy, one needs to know the correct values. Thus, the accuracy of the model is assessed when no intervention occurred – a placebo test. Since there is no energy efficiency impact, any estimated difference between the baseline and actual energy use is, in fact, error. We assessed accuracy both during the onset of COVID (March 19, 2020 through December 31, 2020) and by using back-casting.

For the COVID assessment, we used OBF-AP future 2021 participants because they did not receive an energy efficiency intervention during the training (2019) or out-of-sample testing period (2020). To perform this test, we:

1. Selected 70 electric customers who enrolled in the OBF-AP program after the onset of COVID (and therefore their loads were not affected by energy efficiency during the initial COVID lock-downs that occurred between March and December of 2020).
2. Trained the models using six months of pre-COVID usage data (2019).
3. Predicted customer energy usage during the period of COVID lockdowns.

In addition, we assessed how each of the approaches performed using actual participants by back-casting. The energy model was developed using the year immediately before the energy efficiency installation and back-cast, out-of-sample, to predict a baseline for year prior. For example, if a customer installed energy efficiency in January 1, 2020, data from 2019 was used to develop the energy model, and the model was then used to predict energy usage in 2018. This assessment did not account for the impact of COVID, but helped us select a model that was best-suited to the 2019 participants, as the usage for 2019 participants is highly concentrated among a few customers.

To identify the best performing approach, we used metrics for bias and precision. The primary metric for bias was percent bias, which reflect the percentage by which a method tends to over or underestimate energy use. The metric can be positive or negative and the closer it is to zero the better. The main metric used for precision was the relative root mean squared error. Simply put, the metric measures the relatively magnitude of errors across individual hours (8760), weighting more extreme errors more heavily. Both metrics were measured out-of-sample, meaning that the models are not trained using the time periods used to assess model performance. The metrics were also calculated for the program as a whole (versus at the individual customer level).

Pre-Post NMEC Methodology– No Control Group

Given the impact the pandemic had on non-residential loads, a simple pre-post methodology without control groups does not adequately control for exogenous effects. We used OBF-AP future 2021 participants to assess accuracy because they did not participate in energy during the training (2019) or out-of-sample testing period (2020).

Figure 5 uses the calendar year of 2019 to train the model (“Training Period”) and then predicts on 2020 (“COVID”). The pre-post model incorrectly estimates savings of 6.98% in 2020 even though there is no energy efficiency effect. Without controls, the pre-post methodology assumes that the only change between the pre-period and the post-period is weather and the installation of the efficiency, but it is obvious that this is not the case. While there can be other exogenous factors that impact a customer’s energy use, the over-prediction of savings in this case is a direct result of the COVID pandemic.

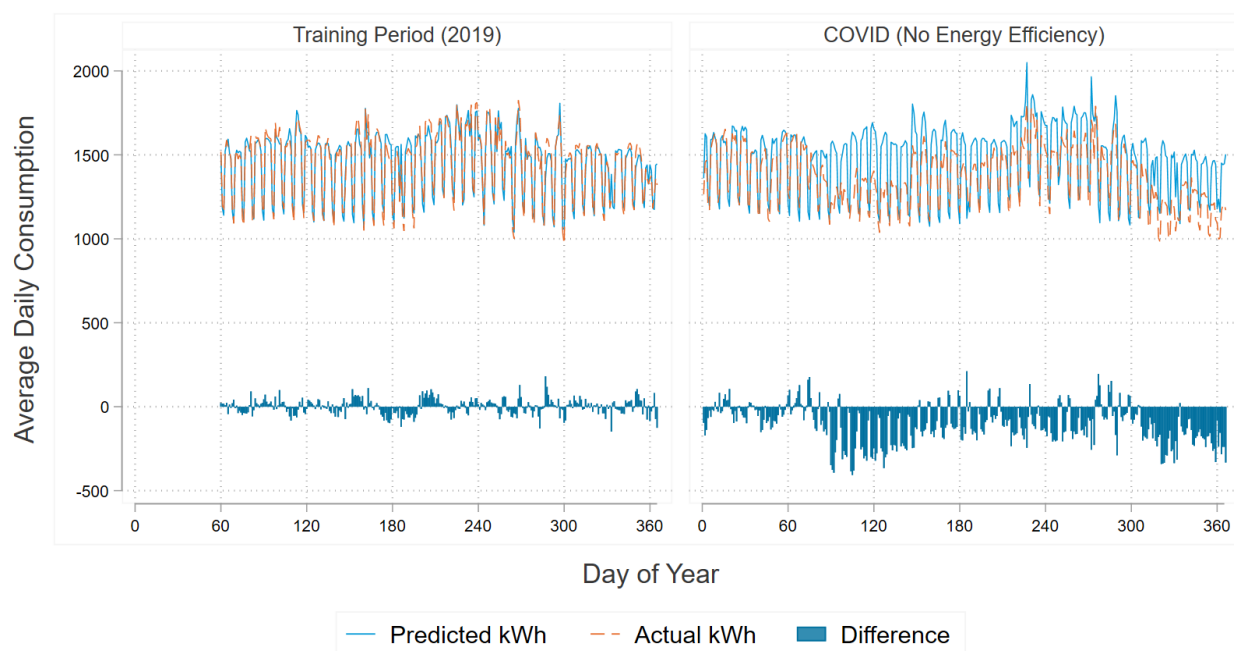


Figure 5: COVID Accuracy for 2021 Electric Participants – No Controls

Performance of Aggregated Controls and Individual Level Controls

In order to determine whether control group methods were robust to non-routine events such as COVID, our team first conducted an accuracy assessment using future 2021 participants. Table 3 summarizes the out of sample bias and precision metrics for each method during the onset of COVID. The model with no controls is included as a reference point, followed by five methods that incorporated control groups into the analysis. As noted above, we can see that the pre-post method with no comparison group tends to over-predict savings by 6.98%. The bias decreases once any of the comparison groups are introduced. The propensity-score matching control group selection method reduced bias the least, while aggregated controls and a stratified random sample control group fared better. Both control candidate pools - the 50,000 non-participants and former energy efficiency participant - were effective at reducing bias.

Table 3: COVID Electric Out-of-Sample Testing Results by Method

Method	N	RRMSE	% Bias
Pre-Post Without Controls	70	0.099	6.98%
Aggregated Controls	70	0.039	2.91%
Non-Participant Stratified Random Sample	70	0.051	2.58%
Non-Participant Propensity Score Matching	70	0.059	4.46%
Former EE Stratified Random Sample	70	0.059	0.41%
Former EE Propensity Score Matching	70	0.094	-4.26%

To illustrate the reduction in bias, Figure 6, just as with Figure 5, uses the calendar year of 2019 to train the model (“Training Period”) and then predicts on 2020 (“COVID”), but uses the aggregated controls method rather than the simple pre-post model. If we compare the two figures, we can see that the magnitude of the error decreases and the bias of the model is reduced from 6.98% to 2.91%. As a result, we can see that we are no longer consistently over- or under-predicting usage during the COVID period.

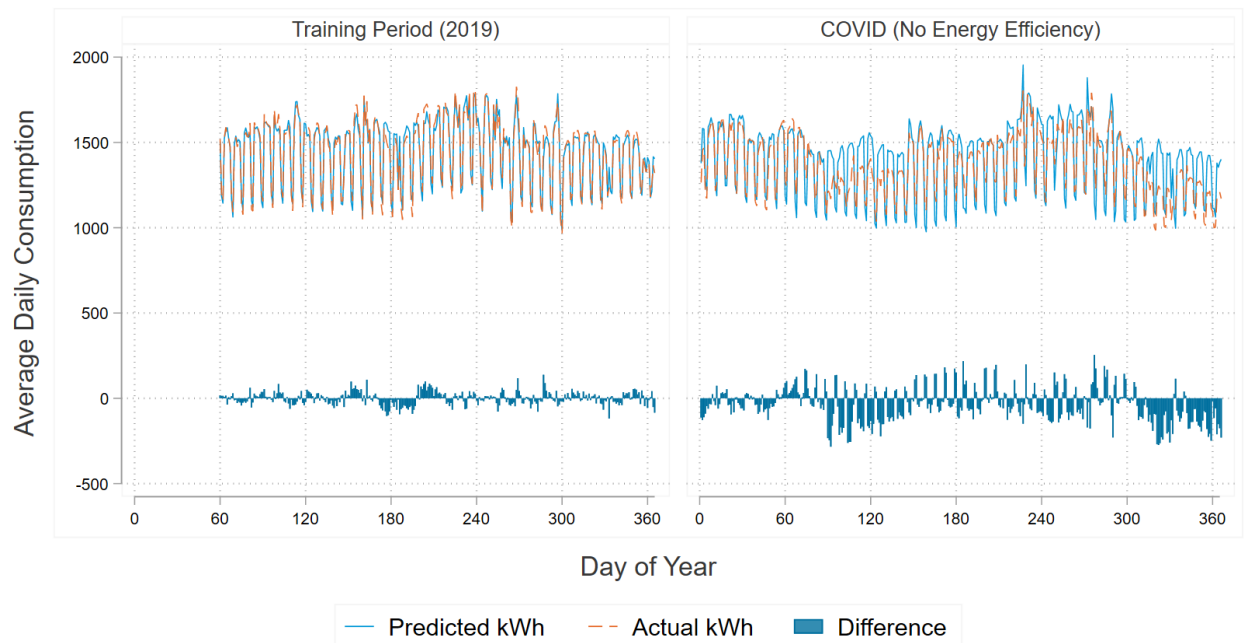


Figure 6: COVID Accuracy for 2021 Electric Participants – Aggregated Controls

Once we determined that all control group methods were able to account for the effect of COVID, we also assessed the accuracy of the models using out-of-sample back casting to select the method best-suited to the 2019 OBF-AP population. As noted previously, the model was trained using each participant’s pre-treatment data (the same pre-treatment data used to estimate energy efficiency impacts) and used to predict electricity usage out-of-sample for 2018, when there was no energy efficiency effect.

Table 4 summarizes the accuracy and precision for the back-casting out-of-sample analysis for each method. The aggregated controls method had the lowest overall bias and the highest overall precision. The difference-in-difference approach with the non-participant stratified-random-sample control group method also performed relatively well, both during COVID and in the back-casting out-of-sample test. While the former energy efficiency participants did relatively well during COVID, they were far less precise for the OBF-AP 2019 participants. Given the success of the aggregated controls in the COVID and the back-casting out-of-sample analysis, this model was selected for estimating participant savings.

Table 4: Back-Casting Out-of-Sample Results by Method

Method	N	RRMSE	% Bias
Pre-Post Without Controls	139	0.047 	-2.02% 
Aggregated Controls	139	0.036 	-1.55% 
Non-Participant Stratified Random Sample	139	0.056 	-2.25% 
Non-Participant Propensity Score Matching	139	0.230 	6.40% 
Former EE Stratified Random Sample	139	0.088 	6.81% 
Former EE Propensity Score Matching	139	0.106 	-5.42% 

Estimated Savings

Figure 7 depicts the model performance of the aggregated control method for electric consumption of 2019 participants. The first panel shows the model performance during the out-of-sample pre-treatment period, the second panel shows the training period, and the third panel shows the savings estimated by the model during the post period. We can see that the model is fairly accurate in the out-of-sample period relative to the training period and does not consistently over- or under-predict usage. We can also see that the error for the model is small relative to the overall savings we see in the post period.

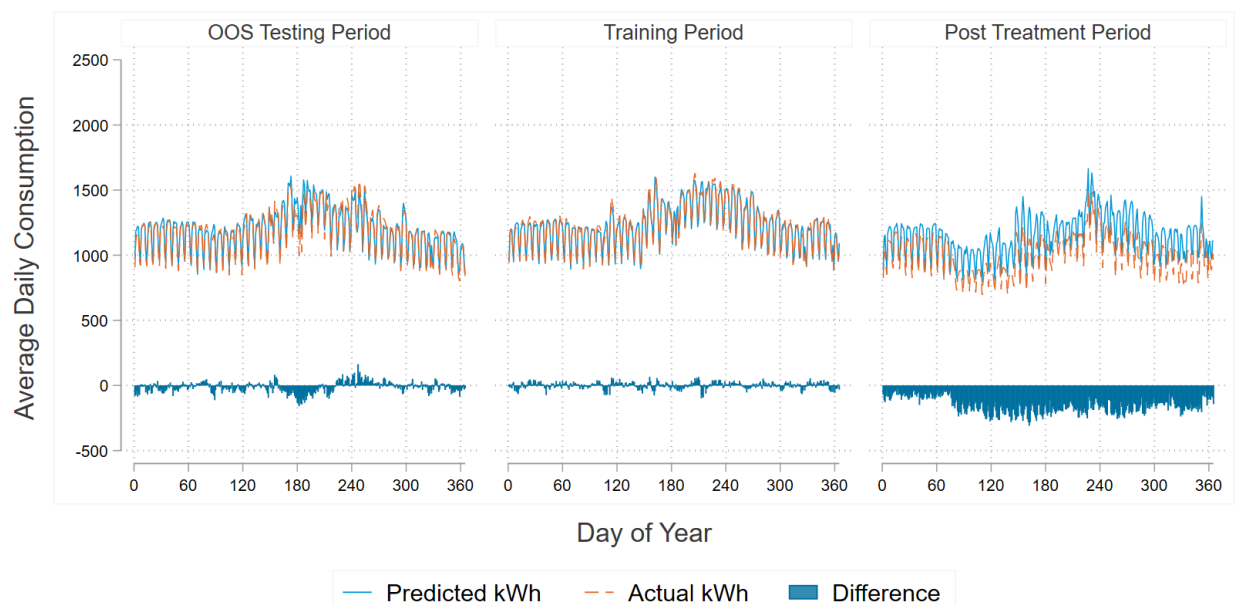


Figure 7: Back-casting Out-of-Sample Accuracy for 2019 Electric Participants – Aggregated Controls

Table 5 depicts the COVID-adjusted first year electric savings for all customers enrolled in OBF-AP during 2019. In total, 139 customers received EE measure installations during 2019, producing 7.37 GWh of savings. Non-lighting projects had the highest percent savings (23%) and contributed the most to the overall portfolio savings despite having only 17 participants. A total of 117 customers received lighting measures and on average saw savings of 8%. This is different from the expected savings, which predicted higher percent savings for customers who received only lighting measures.

Table 5: Electric Energy Savings of 2019 Participants – With COVID Adjustment

Project Type	Participants	Average Annual Predicted Consumption (kWh)	Average Annual Actual Consumption (kWh)	Per Customer Impact (kWh)	Percent Impact	Energy Savings (GWh)
Lighting	117	252,168	231,279	20,889	8%	2.44
Non-Lighting	17	1,190,091	919,844	270,248	23%	4.59
Both	5	1,097,610	1,031,434	66,176	6%	0.33
Total/Average	139	18,272	15,702	2,571	14%	7.37

Conclusions/Key Takeaways

1. **Methods that do not include control groups do not account for the effects of the pandemic.** Models that do not incorporate a control group tend to under-predict customer usage during COVID for OBF-AP participants. The magnitude and direction of the error

varies by segment and interacts with the extreme weather that occurred in 2020 in California.

2. **Comparison groups improve accuracy.** In essentially all cases, the inclusion of a control group, regardless of how it was incorporated into the counterfactual, improved accuracy and precision of the estimates.
3. **Aggregated controls performed better than the difference-in-differences methods that were also tested for the analysis, but there were difference-in-difference methods that were effective.** Aggregated controls performed better than the other models for this study and for this group of customers. However, other studies have found that difference-in-difference methods are also viable options for analysis.
4. **Metrics of accuracy (bias) and precision should be calculated using out-of-sample validation.** As noted above, there are a variety of methods that can be used to predict non-residential energy use. The final method used for this type of analysis should be selected based upon their accuracy and precision, as these metrics measure the tendency to over or underestimate the baseline as well as how close individual hourly or daily estimates are to the actual answer and measure noise. Out-of-sample validation is critical since models that are over-fitted can perform well in sample and poorly out of sample.
5. **There are aspects of estimating customer energy savings that require further study.** In particular, it will be important to assess whether customer savings change over time as the impacts of COVID continue to change.
6. **Methods must be transparent and replicable by all interested parties.** Reproducibility means obtaining consistent computational results using the same input data, computational steps, methods, and code. With the current California laws and regulations that exist to protect the privacy of AMI/smart meter data for individual customers, the use of granular profiles like the ones used in the “aggregated control” method allow for independent review without divulging the specific characteristics of individual participants.