Cleaning Up the Mess of Energy Billing Data: An investigation of Differences in Billing Analysis Results Caused by Data Cleaning Methodologies

Eric Ziemba, Opinion Dynamics, Boston, MA
Stefanie Wayland, Opinion Dynamics, Boston, MA
Olivia Patterson, Opinion Dynamics, Boston, MA

ABSTRACT

A common analysis for energy efficiency program Evaluation, Measurement, & Verification (EM&V), known as a billing analysis, utilizes customer billing records to determine the average energy savings attributable to a utility's energy efficiency program. Typically, billing analyses are used when evaluating programs such as: home energy audits (including low-income variations), and behavioral programs. This paper focuses on home energy audit programs, though the research could be expanded to include other programs where billing analyses are used. Behavioral changes, and additional energy saving upgrades, due to the educational components of these programs cannot be calculated based on data collected during the audit. This leads us towards billing analyses, which generally utilize regression analyses to determine the change in energy consumption caused by equipment upgrades and changes in energy use behavior, brought about though participation in the program.

Importantly, there are a few core elements to billing analysis approaches – 1) data cleaning and preparation, 2) selecting a comparison group, and 3) model specification and validation. This paper focuses on the influence of data cleaning and preparation related to billing analysis results, specifically the influence related to aligning billing periods.

For this paper, we test two period assignment methodologies to maximize the reliability of our models to investigate how data cleaning effects the accuracy of savings estimates. The first method assigns the billing period to the month in which the most days of bill occurs. The second, referred to as calendarization, takes the average daily energy consumption from each bill, assigns that value to each day, and constructs new billing periods based on calendar months.

We find evidence to suggest that how data are cleaned can affect the savings estimates of a billing analysis model, and that, compared to raw data, even small amounts of cleaning may significantly improve the dependability of modeled savings. Additional research is needed, to provide more detail on each of the major data cleaning steps, and which steps and specific methodologies have the most substantial impacts to the reliability of savings estimates.

Introduction

This paper may ask more questions than it answers. However, in the quest to find efficiencies between improved accuracy, and more streamlined processes in our evaluations, a critical first step towards the most accurate answer is to ask the right question. Between the initial findings we present, and our discussion of the data cleaning process for billing analyses, we hope to spark a dynamic and collaborative effort to understand the variation in results associated with moving from dirty to clean data and bring us closer as an industry to finding reliable savings estimates in the most efficient way possible. We hypothesize that data cleaning has a substantial effect on regression results, and calendarization may help to limit biases caused by dirty data.

Billing analyses are commonly used to evaluate programs that include home energy audits, where energy savings are only partially dependent on energy efficient equipment being installed. Low-income direct install programs, that have an educational component are certainly also common. Here, for consistency in our analysis, we focus on core home energy audit programs that are available to customers regardless of income. In this type of program, a contractor visits the home to install a series of low-cost measures (CFL bulbs, weather stripping, etc.), and to educate the resident, or homeowner, about how to be more energy efficient in the home. Behavioral changes, and additional upgrades, due to the educational components of the program cannot be calculated based on data collected during the audit.

Poor data quality has the potential to introduce serious biases in the models used in billing analyses. Data often include missing, or misleading, observations. Variation in meter-read dates can cause inconsistencies in individual billing periods. The potential issues caused by data quality issues, or inconsistencies of billing periods, must be addressed through some method of data cleaning.

Preliminary investigations into how data cleaning can effect savings estimates indicated that the 'cleanliness' of the data might change the final outcome of a billing analysis model. More so, these investigations suggest that methods used while cleaning data, which certainly are not uniform across evaluations and evaluators, can affect final results. Differences in billing analysis results based on the data cleaning the methods used, is cause for concern. Identifying how changes in data cleaning methods that may affect final savings estimates, could lead to improved accuracy in billing analyses for evaluating energy efficiency programs.

We examine how similar impact estimates are, controlling for models and period assignment methodologies, using three 'versions' of data. We compare raw, and partially cleaned, billing data to the results from data that has been fully cleaned for actual program evaluations. Additionally, we test two methodologies for aligning the billing periods to maximize the accuracy of our models, in an effort to improve the accuracy of estimates for savings attributable to the program. We look to quantify the differences in savings provided by each, and show the advantages and pitfalls to each method. The methodologies are tested using case studies from prior evaluations of home energy audits. Additionally, we beg the question; is spending 40+ hours making a dataset as clean as possible, with efforts made to maintain as many records as possible, an efficient use of time when evaluating the kWh impact associated with energy efficiency programs.

Data used comes from four evaluations completed by Opinion Dynamics. While the specific details of the evaluations are confidential, each of them represents the evaluation of a home-audit type program in utility territories in the north eastern quadrant of the United States between 2014 and 2015.

Overview of Billing Analysis

Billing analyses generally utilize regressions to determine the change in energy consumption caused by equipment upgrades and changes in energy use behavior, brought about though participation in the program. Overall, billing analyses allow evaluators to 1) capture behavioral changes and spillover associated with programs designed to intervene in this manner (e.g., audits and HERs), and 2) identify the actual reduction in consumption on the grid that is attributable to the program. Weather data, in the form of heating and cooling degree days, and monthly fixed effects are commonly used to help correct for changes in energy use that occur naturally through the year.

Billing analyses are subject to significant issues due to poor data quality. These data often include missing, or misleading, observations. Common errors include missing or negative energy usage, billing periods that overlap or have gaps between them, and inconsistencies with how estimated or corrected meter reads are dealt with. Additionally, customer energy bills do not have uniform periods, with meters being read at different times of the month for each customer. The variation in meter-read dates can cause billing periods from the same month to represent a range of different time periods, resulting in potential errors. Bills with insufficient, or incorrect, data are commonly dropped from analyses. However, we must often deal with the inconsistencies of billing periods through some method of data cleaning. The following general steps are included in all, or nearly all, billing analyses:

1. Examine data for potentially duplicative data, and flag those instances
2. Check consistency between the days stated for the billing period and the gap between start and end dates; Check for billing period continuity (are there gaps between bills or overlapping billing periods)
3. Assign consistent billing periods (see further discussion on period assignment methods)
4. Merge weather data in some form (e.g. temperature, CDD and HDD, relative humidity)
5. Generate statistics on energy use and weather, including time-series graphs, to evaluate the equivalency of treatment and comparison groups if used
6. Test and validate potential models

The data used for billing analyses generally includes 1) program tracking data, 2) monthly customer billing data, and 3) weather data. Program tracking data provides the date of participation, and may include data on measure installations, participation in other EE programs, and characteristics of the customers' homes (e.g. home-type, size, and heating fuel). The largest and most significant portion of data is the customer energy bills. Since we use actual customer bills from before and after participation, it is possible to observe how their energy use changed after being exposed to the program. Weather data, are also an important piece of the puzzle, since weather is a major contributor to how a customer consumes energy.

The use of a comparison groups to improve the counterfactual of program treatment are a common and generally beneficial methodology used in billing analyses. The use of a solid comparison group, which displays equivalent energy usage levels and trends, has comparable rates of housing stock, and includes customers that have similar traits, can improve the estimate of the counterfactual of how much energy participants would have used if they had not been exposed the program. It is essential to perform equivalency checks between treatment and comparison group customers, which further stresses the need for consistent and reliable data, and data cleaning processes.

Model selection is an incredibly important step in conducting a billing analysis. Obviously finding a model with the best fit is important, but we must also make sure to pay close attention to properly answering the primary question at hand; how much energy was saved as a result of the program being evaluated? Additionally, certain factors must be controlled for to ensure that the coefficient for the variable indicating the time of treatment does not capture a change in energy use that is not associated with the real effect of the program. Commonly, the means including some form of weather data, and potentially data on the type of heating used (Electric, Gas, etc.), and other home characteristics if available.

Methodology

In this section we review the hypothesis, outline the experimental design, and describe the two types of period assignment methodologies used in our research in more detail. We also take a look at three different versions of data used in the experiment, and the amount of cleaning that was involved in each.

Our hypothesis is that the extent to which data are cleaned, and the methods used, can affect regression results in billing analysis. Cleaning data prior to modeling helps reduce potential biases that could be caused by common issues in raw data, including: Consumption measurement error, long bill periods, missing bills, repeated data, estimated consumption, adjusted consumption, incorrect location/address, and incorrect audit/measure installation date. Errors in recorded energy consumption, bill periods, and audit dates are potential sources for errors in the data that could substantially increase bias. We address errors in recorded consumption by looking for outlying consumption values that do not match the rest of a customer's history, and checking for estimated or adjusted consumption values as flagged by the utility. We check for correct billing periods by looking at dates surrounding each bill to see whether start and end dates align and looking for periods that aren't covered by a bill. Finding incorrect audit dates can be much more challenging and while usually outside the purview of billing data cleaning, requires reviewing utility, implementer and sometimes contractor documents and processes.

While our experiment is not focused on specific issues that arise from data cleaning, future work should expand on the effects of specific data cleaning steps and how each could address certain biases.

The Experiment

In this experiment, we hypothesize that data cleaning has a significant effect on regression results, and calendarization may help to limit biases caused by dirty data. To test this, we run a series of Linear Fixed Effects Regression (LFER) models using data that are raw, data where duplicative records are simply removed, and data that had previously been fully cleaned for actual evaluations. For each of these sets of data, we use two different methodologies for assigning billing periods, calendarization and the mid-point method as described below. Two models are used to improve the validity of our results and allow us to see if results from each set of data are model dependent.

The primary measure for determining whether or not data cleaning effects results is whether or not the result of a given model's energy impacts estimates is within the bounds of the same model that uses the final data. Additional statistics will be investigated and reported to highlight potential differences between the quality of results given data cleanliness and/or period assignment method.

Data Cleaning Steps

We used three versions of data to reflect stages of the data cleaning process (Raw, Deduped, and Final).

The version that we refer to as Raw Data includes data that is exactly as delivered from the utility. Variable names were altered for consistency. The billing start date was generated for one utility, which does not provide it. To generate this date, we took the end date, subtracted the number of billing days listed.

The next version of data, Deduped Data, takes the Raw Data, and removes all duplicated records, or "dedupes". A duplicate record can have exactly the same data as another, in which case we simply remove one. In other cases, multiple records could show the same date with contradictory values for energy usage, which would then require additional investigation into how to determine the most accurate information based on energy trends or records that have been flagged as estimated or adjusted. In this version of data, simplest approach, and drop all records that have any duplicative information. Deduping, to some degree of rigor, is a necessary data cleaning step that, using the statistical program STATA, can be completed relatively quickly. In our case, deduping the data by merely removing all duplicates, took very little time, and eliminated 11.7% of the records from the raw data.

The last version of data, Final Data, is a compilation of the datasets used for the final models in each of the respective evaluations. These data went through detailed and extensive cleaning, with an effort to maintain to most data possible[1]. We use these data as a benchmark for accurate results. In testing the importance of data cleaning and period assignment methodologies, we compare the point estimates of energy savings from the prior versions of data against the error bounds of results from models using the Final Data.

Period Assignment Methods

For this paper we test two methodologies for aligning the billing periods to maximize the accuracy of our models, in an effort to improve the estimates for savings attributable to the program. We look to quantify the differences in savings provided by each, and show the advantages and pitfalls to each method. The methodologies are tested using case studies from prior evaluations of home energy audits, conducted by our firm.

To assign consistent periods in billing data, the Mid-Point method is common. With this method the period is assigned to the month in which the most days of bill occurs. By using this method, the original integrity of the data are maintained. The kWh values for the billing period are the actual reading of energy consumption for the given time period. In seeking the most accurate estimation of program impacts, using the actual kWh reading from a billing period seems logical. As we have mentioned, however, there is often substantial fluctuation in the time of month at which meters are read for each customer. Even for a single customer, the meter read date is not often constant (i.e. always the 1st or 15th day). Due to this inconsistency, assigning the period in this way can lead to a couple potential complications. 1) Periods with the same designation may not be fully comparable, because there is the potential for the same period for two different customers to only represent a small number of the same days. For example: using the mid-point method, a June period for one customer could be a bill from May 18th – June 19th, where another customer's June period could be June 15th – July 14th. These two periods would be compared in the model, despite only having five days in common, causing potential errors. 2) In cases where meters are read in the middle of the month, it is likely that duplicative periods could be assigned to consecutive billing records using this methodology. For instance, bills from February and March could both be assigned to March, if the regular meter read is on the 15th since February

[1] These data had all perfect-duplicates removed. Data were also check thoroughly for inconsistencies in billing dates and energy usage. Adjustments to the dates of billing periods may have been shifted by a limited number of days to adjust for gaps or small overlaps. Overlapping periods are generally combined, and divided equally with billing days and energy use. Significant outliers were also removed.

on has 28 days (See Table 1.). These duplicates would need to be reassigned to consecutive months, or collapsed in to one period containing data from two months. While collapsing the data does not necessary cause significant problems, it does reduce the granularity of the information available, and could cause a similar issue to the first complication discussed, where two months' worth of consumption and weather data would be compared to only one (or two bi-month bills compared that overlap only one month). These complications often do not effect a high percentage of the data, but must be closely checked, increasing the time necessary to fully clean a dataset.

Table 1. Duplicative Period Assignment Examples

Account	Bill Start Date	Bill End Date	Mid-Point	Period	Issue
1001	7/16/2013	8/13/2013	7/30/2013	Jul-13	Only two days in common
1002	6/14/2013	7/18/2013	7/1/2013	Jul-13	
1003	2/18/2015	3/14/2015	3/2/2015	Mar-15	Consecutive periods with read in middle of month
1003	3/15/2015	4/17/2015	3/31/2015	Mar-15	

The other period assignment method that we investigate is calendarization, which takes the average daily energy consumption from each bill, assigns that value to each day, and constructs new billing periods based on calendar months. This methodology is effectively applying a smoothing function to the consumption before modeling, which means that we expect that it will bias error estimates low, and potentially could have additional effects on months at the start and end of the billing records for each customer. Calendarization, depending on the exact method used, could increase or decrease the actual number of data points, which could also have the effect of biasing the error estimates. From a savings impact estimation standpoint, we don't expect that calendarization will have a large effect on the estimate itself, especially if the analyst makes reasonable choices around blackout dates.

Models

When conducting billing analyses for these types of programs, we generally utilize some form of a LFER model to assess the energy savings attributable program. In this research we include two models, to help control for model dependency, and show that results we find are due to data cleaning, or period assignment methodologies. All models that we use include a "future participant" comparison group, which consists of pre-period data for participants from the program year following the one being evaluated. This type of comparison group assumes that comparison group participants are similar to those in the treatment group, because they are known to have an equivalent propensity for participating in a program of this nature. Equivalency checks are undertaken, and outlier accounts are removed to ensure a reasonable counterfactual to treatment in the program.

The first model used is fairly basic, controlling for weather overall, and potential changes in the effect that weather has on energy consumption in the post period. Failing to account for non-program-related changes that occur during the post-participation period, for example, the warmer summers that have been experienced, could undervalue the treatment effect.

Equation 1. Linear Fixed Effects Controlling for Post-Period Weather

$$ADC_{it} = B_h + B_1 Post_{it} + B_2 HDD_{it} + B_3 CDD_{it} + B_4 Post \cdot HDD_{it} + B_5 Post \cdot CDD_{it} + \varepsilon_{it}$$

The second model builds on the first, and controls for seasonal changes in energy use, through the inclusion of terms for each month of the year (January–December). This allows a month to be present in both the pre-participation period and the post-participation period, thus capturing the change in usage during said month. Our use of these monthly terms in conjunction with a comparison group creates an improved counterfactual and increases the accuracy of program savings estimates.

To adjust for differences in pre-participation period energy use between our treatment and comparison groups, this model includes interactions of the treatment with monthly terms to control for those inconsistencies. We also interact the effects of each month with the post-participation period, to control for changes in how seasonality affects energy consumption.

Equation 2. Two-Way Fixed Effects Model

$$ADC_{it} = B_h + B_1Post_{it} + B_2HDD_{it} + B_3CDD_{it} + B_4Post \cdot HDD_{it} + B_5Post \cdot CDD_{it} + B_tMonth + B_{t1}Month \cdot Post + B_{t2}Month \cdot Treat + \varepsilon_{it}$$

Results

In our investigation of data cleaning steps we looked at results from models using each version of the data, each with two methods for assigning billing periods. Controlling for two model specifications and two period assignment methodologies, nearly half (43%) of the results using raw data fell within the error bounds of the results from data that had been fully cleaned during the evaluation. In Table 2 and Table 3, we show the count of models that had a savings estimate that was within the error bounds of results from the same program's Final Data. Additionally, we include a mean "scaled value" that represents how close, on average, the estimates were. This scale was calculated as the difference between the two results, over ½ error bounds. A value of 1 would indicate that the results are exactly on the edge of the error bounds. Any value less than 1, approaching 0, is increasingly close to the final savings estimate. Values over one indicate increased distance from the final estimate.

Equation 3. Scaled Value of Estimate Accuracy

$$\text{Scaled Value} = \frac{|Estimate - Final|}{Error}$$

Table 2. Results from Modeling Using Raw Data

Raw results within Error-Bounds of Final	Calendarized		Mid-Point		Total	
	Count	Mean Scaled Value	Count	Mean Scaled Value	Count	Mean Scaled Value
Yes	4	0.16	3	0.68	7	0.38
No	4	2.05	5	3.60	9	2.91
Total	8	1.11	8	2.51	16	1.81

When making a slight increase to the rigor used in data cleaning, by deduping, the rate at which the results fall within the error bounds of final results increases to 63%. We also can see that, overall, the mean scaled value falls just below 1. The improved scale value is mostly driven by the models using Mid-Point data. The Mid-Point data show an increased rate of modeled results within the error-bounds of final estimate, as we had expected. It is interesting that calendarization maintains the same count of models within the final error-bounds, but with slightly lower accuracy for those results within.

Table 3. Results from Modeling Using Deduped Data

Deduped results within Error-Bounds of Final	Calendarized		Mid-Point		Total	
	Count	Mean Scaled Value	Count	Mean Scaled Value	Count	Mean Scaled Value
Yes	4	0.31	6	0.48	10	0.41
No	4	1.96	2	1.94	6	1.96
Total	8	1.14	8	0.84	16	0.99

Do we need to clean as much as we do? Maybe not, given that simply deduped data yielded significant improvements to the overall rate of results within the error-bounds of final savings estimates, and the accuracy of those results. However, we certainly would not want to stop here, with the cleaning process. However, this provides evidence that some individual data cleaning steps could provide substantial improvement to the accuracy. Variation in how cleaning steps impact model results could, in part, be a function of the quality of the original data. However, given that this study uses data from four different programs, each with their own unique data issues, some degree of the initial cleanliness of data is controlled for. With what we how found here, e may go so far as to postulate that a more uniform and simplified, or streamlined, data cleaning process could be advantageous, and provide reliable results without spending so much time. For the moment, there is considerable conjecture in the theory, and further investigation is necessary and important. Also, at what rate of results being with error bounds would we want to achieve to determine the most efficient amount of data cleaning necessary? If pressed for time, can certain steps be bypassed, and would focusing on certain cleaning steps clean the data in such a way that results can be relied upon by clients? These are all important questions that need to be answered, and could be answered by expanding on the type of research presented here.

An additional, and potentially important, finding relates to the R-Squared values from models that were run using each of the period assignment methodologies. Models using calendarized billing data result in significantly higher R-squared values across all datasets used. This could be a product of the smoothing that occurs with calendarization. However, given the limited scope of this study, it is difficult to draw concrete conclusions.

Table 4. Statistics on R-Squared Values from Models Using Each Dataset

Method	Statistic	Raw Data	Deduped Data	Final Data
Calendarized	Min	0.597	0.61	0.613
	Mean	0.648	0.668	0.675
	Max	0.718	0.718	0.724
Mid-Point	Min	0.533	0.581	0.585
	Mean	0.576	0.625	0.639
	Max	0.635	0.674	0.689
Total	Min	0.533	0.581	0.585
	Mean	0.612	0.646	0.657
	Max	0.718	0.718	0.724

Overall, the results we show here suggest two key things. First, with raw data, calendarizing billing data may provide more robust and accurate savings estimates than re-assigning periods based on the mid-point of the billing periods provided. Second, on the whole, it appears that deduping raw data could, at the very least, substantially improve the accuracy of savings estimates. We do not yet have a solid understanding of why deduped data that has been calendarized does not provide improved accuracy, however, given the limited scope of this study, we chose to delay making conclusions on whether or not calendarization of billing data mitigates biases associated with dirty data. Additional research should be done, to provide more detail on the effects of each of the major data cleaning steps, and how calendarization impacts those effects.

Discussion

We all know about the importance of having reliable data when conducting any type of statistical or quantitative analysis. It is crucial to both the quality of our work, our confidence in results provided to clients, and quite often our own sanity, that the data we work with can be trusted. Errors in data collection, whether due to equipment malfunction, or simple human error, often cannot be helped. Any researcher who handles data must be aware of potential data issues. Fault cannot, and should not, be assigned for the normal maladies that occur in collected data. It is our job to verify the integrity of the data we receive, and take steps to appropriately address any issues that we find.

In the case of energy billing data, we trust in the validity of the data provided our utility clients, and continuously work to ensure that all records included represent the clearest picture of reality. At times we make adjustments to dates if there are questions surrounding the accuracy of that information. Other times we must correct an estimated energy use by incorporating information from other relevant records that are included to show the final usage or an adjustment that was made after the initial meter-read. In general though, most of the data provided to us is accurate. Across the data used in this study, upwards of 90% of records in each dataset did not require any significant cleaning. The records that do require cleaning, still provide us with important and useful information, if at the cost of some effort and maybe a new gray hair. If left unclean, these data could skew the estimates of program savings that we are investigating. Alternatively, simply removing them, could leave us with an incomplete picture of how a program affects a customer's energy use.

The results of this study, provide some evidence to suggest that rigorous cleaning may not be hugely important, and that even potentially simple solutions to handling dirty data may play a significant role in improving the validity of the data being used for modeling (and the results of those models). As Advanced Metering Infrastructure (AMI) becomes more prevalent, issues regarding errors in data reads may be less severe due to the increased granularity of the data. More work on this topic is imperative to improving our understanding of the effects that specific data cleaning steps, and the methodologies utilized when preparing data for analysis. We plan to continue along the lines of this study, to include additional standard cleaning steps, and period assignment methodologies. We also wish to investigate the impacts of how outliers are dealt with, and the criteria for including or removing specific accounts from the final models. In addition to our work, we hope to see more evaluators take part in studying how their standards for data cleaning affect the results of billing analyses. Future research could potentially use simulations to determine which cleaning steps provide the most "bang for the buck", if budgets are tight. It could be useful for both evaluators and our clients to learn which cleaning steps will get you closest to reliable results if, for instance, you only have a day (instead of a week) to spend on cleaning data.

Why is this important, aside from the desire to be rigorous in our research, or the need to give our best work to our clients? Data cleaning is frequently a "black box" to our clients, and sometimes even colleagues working on other components of an evaluation. This lack of transparency isn't really benefiting anyone. Clients often have few ways to check that data have been cleaned sufficiently and without bias. Improving our data cleaning processes, so as not to spend too much time on steps that may not impact the validity of data, could cut down on project costs or leave more time for deeper exploration into how a given program impacts energy use. Our industry lacks uniformity in our methods for cleaning data, and as we can see from the findings of this preliminary study, how data are cleaned can change the final result. This makes year to year comparisons difficult, and potentially unreliable in cases where a client hires a new evaluator. Developing standards, which research of this type could help to inform, could improve the reliability of program savings estimates from billing analyses across our industry. Certainly, we all have unique ways of answering the same question, and arriving at the same answer. However, having clear guidelines for cleaning billing data while understanding how cleaning steps and methods can affect results, would streamline our analyses and benefit all parties involved.

How do you clean data? What steps do you believe are most important? Must data be perfect to deliver results that can be relied upon by clients to make important decisions about the future of their programs? Where can evaluators make improvements to their data cleaning processes? What really drives differences in the results from models that use raw, clean, or partially-clean data? All of these questions are important to answer. We will continue our efforts to uncover evidence to answer them. Will you do the same?

Appendix

Definition of Terms in Model Equations

Equation 1. Linear Fixed Effects Controlling for Post-Period Weather

$$ADC_{it} = B_h + B_1 Post_{it} + B_2 HDD_{it} + B_3 CDD_{it} + B_4 Post \cdot HDD_{it} + B_5 Post \cdot CDD_{it} + \varepsilon_{it}$$

ADC_{it}	= Average daily consumption (in kWh) for the billing period
$Post$	= Indicator for treatment group in post-participation period (coded "0" if treatment group in pre-participation period or comparison group in all periods, coded "1" in post-participation period for treatment group)
HDD	= Average daily heating degree days from NCDC
CDD	= Average daily cooling degree days from NCDC
B_h	= Average household-specific constant
B_1	= Main program effect (change in ADC associated with being a participant in the post-program period)
B_2	= Change in ADC associated with one unit increase in HDD
B_3	= Change in ADC associated with one unit increase in CDD
B_4	= Change in ADC associated with each increment increase of HDD for participants in the post-program period (the additional program effect due to HDD)
B_5	= Change in ADC associated with each increment increase of CDD for participants in the post-program period (the additional program effect due to CDD)
B_t	= Coefficients for each month
ε_{it}	= Error term

Equation 2. Two-Way Fixed Effects Model

$$ADC_{it} = B_h + B_1 Post_{it} + B_2 HDD_{it} + B_3 CDD_{it} + B_4 Post \cdot HDD_{it} + B_5 Post \cdot CDD_{it} + B_t Month + B_{t1} Month \cdot Post + B_{t2} Month \cdot Treat + \varepsilon_{it}$$

ADC_{it}	= Average daily consumption (in kWh) for the billing period
$Post$	= Indicator for treatment group in post-participation period (coded "0" if treatment group in pre-participation period or comparison group in all periods, coded "1" in post-participation period for treatment group)
HDD	= Average daily heating degree days from NCDC
CDD	= Average daily cooling degree days from NCDC
$Month$	= Month indicator
$Treat$	= Indicator for treatment group participants
B_h	= Average household-specific constant
B_1	= Main program effect (change in ADC associated with being a participant in the post-program period)
B_2	= Change in ADC associated with one unit increase in HDD
B_3	= Change in ADC associated with one unit increase in CDD
B_4	= Change in ADC associated with each increment increase of HDD for participants in the post-program period (the additional program effect due to HDD)
B_5	= Change in ADC associated with each increment increase of CDD for participants in the post-program period (the additional program effect due to CDD)
B_t	= Coefficients for each month
B_{t1}	= Coefficients for each month in the post-participation period
B_{t2}	= Coefficients for each month for treatment group participants
ε_{it}	= Error term

Model Results Summary

The following table shows the final savings estimate, error bounds, and key statistics, from each model.

		Program 3				Program 4			
	Model	Model A	Model B	Model A	Model B	Model A	Model B	Model A	Model B
	Period Assignment	Calendarized	Calendarized	Mid-Point	Mid-Point	Calendarized	Calendarized	Mid-Point	Mid-Point
Raw Data	Daily Savings Estimate (kWh)	-2.329	-2.982	-2.699	-3.078	-4.143	-2.759	-5.010	-4.079
	Lower Bound (90% CI)	-2.655	-3.324	-3.197	-3.639	-4.497	-3.109	-5.514	-4.574
	Upper Bound (90% CI)	-2.003	-2.641	-2.201	-2.518	-3.789	-2.409	-4.507	-3.583
	R-Squared	0.659	0.664	0.537	0.554	0.701	0.718	0.622	0.635
	Std. Deviation	0.198	0.207	0.303	0.341	0.215	0.213	0.306	0.301
	Precision	14%	11%	18%	18%	9%	13%	10%	12%
Deduped Data	Daily Savings Estimate (kWh)	-2.317	-2.956	-2.384	-2.693	-4.143	-2.759	-5.010	-4.079
	Lower Bound (90% CI)	-2.635	-3.289	-2.745	-3.083	-4.497	-3.109	-5.514	-4.574
	Upper Bound (90% CI)	-1.998	-2.622	-2.024	-2.303	-3.789	-2.409	-4.507	-3.583
	R-Squared	0.711	0.716	0.669	0.674	0.701	0.718	0.622	0.635
	Std. Deviation	0.194	0.203	0.219	0.237	0.215	0.213	0.306	0.301
	Precision	14%	11%	15%	14%	9%	13%	10%	12%
Final Data	Daily Savings Estimate (kWh)	-1.563	-2.249	-1.804	-1.869	-4.116	-2.763	-4.522	-3.814
	Lower Bound (90% CI)	-1.857	-2.555	-2.154	-2.239	-4.466	-3.110	-5.091	-4.408
	Upper Bound (90% CI)	-1.268	-1.944	-1.455	-1.499	-3.765	-2.416	-3.952	-3.219
	R-Squared	0.716	0.723	0.685	0.689	0.708	0.724	0.666	0.680
	Std. Deviation	0.179	0.186	0.212	0.225	0.213	0.211	0.346	0.362
	Precision	19%	14%	19%	20%	9%	13%	13%	16%
	Raw results within Error-Bounds of Final	0	0	0	0	1	1	1	1
	Deduped results within Error-Bounds of Final	0	0	0	0	1	1	1	1

		Program 1				Program 2			
	Model	Model A	Model B	Model A	Model B	Model A	Model B	Model A	Model B
	Period Assignment	Calendarized	Calendarized	Mid-Point	Mid-Point	Calendarized	Calendarized	Mid-Point	Mid-Point
Raw Data	Daily Savings Estimate (kWh)	-3.885	-5.241	-5.326	-6.885	-1.471	-1.969	-1.752	-1.668
	Lower Bound (90% CI)	-4.406	-5.752	-5.835	-7.389	-1.797	-2.300	-2.090	-1.997
	Upper Bound (90% CI)	-3.365	-4.731	-4.817	-6.381	-1.145	-1.639	-1.414	-1.339
	R-Squared	0.602	0.620	0.539	0.589	0.597	0.621	0.533	0.595
	Std. Deviation	0.316	0.310	0.310	0.306	0.198	0.201	0.206	0.200
	Precision	13%	10%	10%	7%	22%	17%	19%	20%
Deduped Data	Daily Savings Estimate (kWh)	-3.476	-4.886	-3.454	-4.310	-1.954	-2.093	-2.005	-2.119
	Lower Bound (90% CI)	-3.980	-5.372	-3.931	-4.783	-2.288	-2.428	-2.350	-2.462
	Upper Bound (90% CI)	-2.972	-4.400	-2.977	-3.838	-1.621	-1.758	-1.660	-1.777
	R-Squared	0.610	0.628	0.593	0.609	0.616	0.640	0.581	0.613
	Std. Deviation	0.306	0.296	0.290	0.287	0.203	0.204	0.210	0.208
	Precision	14%	10%	14%	11%	17%	16%	17%	16%
Final Data	Daily Savings Estimate (kWh)	-3.202	-4.342	-3.306	-4.145	-1.369	-1.895	-1.975	-2.376
	Lower Bound (90% CI)	-3.707	-4.827	-3.790	-4.621	-1.683	-2.219	-2.281	-2.690
	Upper Bound (90% CI)	-2.696	-3.856	-2.823	-3.669	-1.054	-1.572	-1.669	-2.061
	R-Squared	0.613	0.628	0.595	0.608	0.638	0.650	0.585	0.601
	Std. Deviation	0.307	0.295	0.294	0.289	0.191	0.197	0.186	0.191
	Precision	16%	11%	15%	11%	23%	17%	15%	13%
	Raw results within Error-Bounds of Final	0	0	0	0	1	1	1	0
	Deduped results within Error-Bounds of Final	1	0	1	1	0	1	1	1