Estimating Energy Savings Resulting from Strategic Energy Management Programs: Regression Framework Comparison

Andrew Bernath, The Cadmus Group, Portland, OR
Maggie Buffum, The Cadmus Group, Portland, OR

ABSTRACT

Strategic Energy Management (SEM) programs seek to implement long-lasting, comprehensive changes in energy management and consumption at industrial facilities.[1] Evaluators typically estimate energy savings associated with SEM programs using regression analysis to model facility energy consumption as a function of weather and other variables, comparing energy consumption before and after program implementation. There are several regression model frameworks to choose from, and selecting the right framework is integral to completing an accurate, robust evaluation. Testing frameworks to determine which is best suited to evaluate savings for a particular facility (and under a specific set of circumstances) presents a challenge because the true energy savings value at a facility is unknown. However, using simulated data, we can start to answer questions about which regression models perform best under various circumstances because we know the "true" savings at the facility. In this study, we simulated energy consumption data representing a SEM program implementation at a typical industrial facility to answer two important questions: Which regression analysis framework should evaluators use to calculate accurate, robust savings estimates? How should the facility regression model be specified within that framework?

Summary of Findings

The simulation revealed several important findings:
- The fully specified pre/post model framework produces unbiased estimates and captures the true savings at the nominal 80% confidence level under more scenarios than the other two frameworks for both simple and complex facility specifications.
- For facilities that have simple model specifications, the simple pre/post model framework performs comparably to the forecast and fully specified pre/post models. For facilities with complex model specifications, the simple pre/post framework was unreliable, consistently producing biased savings and failing to reach the nominal capture rates.
- Nearly all of the regression frameworks produced biased estimates, increased standard errors, and poor capture rates under scenarios where key variables were omitted from the model.
- In the scenario where an event affecting energy consumption occurs during the post period, an indicator variable can be included in a pre/post model to result in unbiased savings estimates and high capture rates. However, without an engineering estimate available to adjust savings in a forecast model, the forecast model produces biased savings estimates with low capture rates.
- On average, savings were unbiased for the scenario in which the regression model specification did not account for autocorrelation, however, standard errors of savings estimates were increased and the capture rates for the true savings were poor.

---

[1] SEM is implemented in commercial facilities as well. The focus of this research is on industrial SEM evaluation.

Introduction

Strategic Energy Management (SEM) are a rapidly growing approach to energy efficiency. SEM programs use capital and operations and management (O&M) improvements to facilities to realize energy savings. Various frameworks are used to estimate reductions in energy consumption before and after the programs are implemented. They typically involve facility-level regression models but use different model specifications and underlying assumptions. Industrial facility models pose challenges due to the large numbers of variables that effect energy consumption and the relatively little data we have to capture those relationships. They typically include complex interactions between drivers of energy consumption and sometimes include non-routine adjustments before or after the SEM program begins, which impact energy consumption but are not related to the program. These non-routine adjustments may or may not have engineering estimates of energy savings. Furthermore, determining which regression framework and model specification to use in an evaluation is challenging because there is no industry-standard that produces results with proven accuracy under a variety of possible scenarios.

In this study, we used a simulation approach to test the different frameworks and model specifications. Our goal was to shed light on which ones produce accurate, robust results and under what conditions, providing guidance to future evaluations. A simulation approach is particularly effective for this purpose because we determined what the "true" savings are and model facility energy consumption based on this value of savings as well as other variables that drive energy consumption. Because we know the savings underlying the simulated data, we can compare estimates resulting from different regression model frameworks to the truth to determine how accurate the estimates are, on average. We assessed the accuracy and robustness of the following three regression frameworks, described in detail below:

- Forecast
- Simple Pre/Post
- Fully-specified Pre/Post

Further, in our evaluations of actual facility data, we have observed that results tended to vary greatly depending on weather and facility conditions, including the following:

- Weather (HDD, CDD, mean temperature)
- Production and occupancy
- Weekday versus weekend activity
- Shutdown or closure periods
- Non-routine adjustments

Finally, due to differing data collection protocols and procedures, evaluators may not have access to data on certain variables. Some data may exhibit autocorrelation or non-normality that aren't included in the regression model. Therefore, we considered evaluation when model specification includes and does not include the following:

- Include extraneous variables or omit known drivers of energy consumption
- Correctly specify or mis-specify the relationships between variables and energy consumption
- Include or not include non-routine adjustments in the regression model
- Account for or do not account for time-series autocorrelation or non-normality in the regression analysis

In the remainder of this paper, we provide details on the regression frameworks, weather and facility conditions, and additional considerations that we examined using the simulation study. We

summarize the findings to identify trends in how accurately each framework estimates energy savings under different circumstances. Ultimately, we provide answers to two important questions:

- Which regression analysis framework should evaluators use to calculate accurate savings estimates?
- How should the facility regression model be specified within that framework?

Methodology

Evaluators typically estimate SEM program energy savings by comparing energy consumption prior to the SEM program (baseline) to energy consumption after the SEM program was implemented. Baseline models are built using only pre-program period data and represent energy consumption at the facility in the absence of the program. Evaluators typically require data on several possible energy drivers to model the baseline, including:

- Weather (HDD, CDD, mean temperature)
- Production and occupancy
- Weekday versus weekend activity
- Shutdown or closure periods
- Timing of non-routine adjustments

In many facilities, some or all of these variables are correlated with energy consumption but in others, they may not be. Using the simulated data, we built several baseline models to test the effects of misspecification.

Once the baseline has been established, the evaluator models post-program period energy consumption to find the difference before and after implementation of the program and calculate savings. We explored three common frameworks for doing this. In one, pre-program energy consumption models are used to forecast energy consumption in the post-program period, but absent the effects of the SEM program. Savings are calculated as the differences between predicted baseline usage and the metered usage. In the second, the baseline model specification is used, with the addition of a post-program indicator. Then savings are estimated based on the coefficient corresponding to the indicator, which represents the average energy savings per time interval. In the third, the baseline model is also used, but effects of the other variables are estimated to capture both pre-program and post-program effects (e.g., the effect of HDD can vary with changes in production or production may interact with program effects). We summarize these frameworks, providing additional details below:

1. **Forecast:** use a baseline regression model to predict what energy consumption would be in the post-program period, absent the program. Sum the differences between predicted usage and metered usage to estimate total savings during the post-program period.
2. **Simple Pre/Post:** use a baseline regression model, specified based on pre-program energy consumption and predictor variables. Estimate the model, with an additional indicator signaling the start of the post-program period. Use the coefficient of the post-program period indicator to estimate average energy savings per time interval (i.e., day, week, month - depending on data frequency, one or more of these may be an option). Multiply the average by the number of time intervals in the post-program period to estimate total energy savings during the post-program period.
3. **Fully-specified Pre/Post:** similar to the simple pre/post framework, use a baseline regression model but in this framework, include interaction terms with the post-program indicator and all

predictor variables. The interactions allow the predictors to have different effects on energy consumption in the pre- and post-program periods. Estimate the model, take the coefficient of the post-program period indicator (main effect) multiplied by the number of post-program time periods, add each of the coefficients of the post-program period interactions multiplied by the sums of their respective variable values during the post-program period, and estimate total energy savings during the post-program period.

Simulated Data

We simulated two data sets to represent facilities like those that we have observed in SEM program evaluations. One data set represents a facility with energy consumption as a function of one production process, an indicator of interruptions to production, and weather (specifically HDD, or heating degree days). We call this the "simple facility", where energy consumption is driven by two variables and no interactions. The second data set represents a facility with energy consumption as a function of two different production processes, an indicator of interruptions to production, weather (HDD), the interaction of weather and production, and a non-routine adjustment (which we refer to as an "event") occurring in the pre-program period which results in a reduction in energy consumption. We call this the "complex facility", where energy consumption is driven by a number of variables and there are interacting energy drivers. In both models, we also included a non-routine event in the post-program period which has an engineering estimate associated with the change in energy consumption. Such events frequently occur in industrial facilities. Some examples are installation of new equipment at a facility, temporary or permanent closures of part of the facility, or staffing changes.

The following equations**Error! Reference source not found.** provide the "true" baseline models for the simple and complex facility (pre/post models also include a post-period indicator and interactions between the post-period indicator and each of the other model variables):

Simple Facility: $kWh_t = \beta_0 + \beta_1 Production_t + \beta_2\ Production\ Interruptions_t + \beta_3 HDD_t + \varepsilon_t$

Complex Facility: $kWh_t = \beta_0 + \beta_1 Production1_t + \beta_2 Production2_t$
$$+ \beta_3 Production\ Interruptions_t + \beta_4 HDD_t$$
$$+ \beta_5 HDD_t x Production1_t + \beta_6 Pre\text{-}Program\ Event_t + \varepsilon_t$$

Where:

β_i	= The coefficient of the ith variable in the model (i = 0 represents the model intercept).
kWh_t	= Energy consumption at the facility at time t
$Production_t$	= Production at the facility at time t
$Production\ Interruptions_t$	= An indicator of interruptions to production at time t
HDD_t	= Heating degree days at time t
HDD_t x $Production_t$	= The interaction between HDD and production at time t
$Pre\text{-}Program\ Event_t$	= An indicator representing a non-program related change in energy consumption at the facility at time t

We specified the error term using random draws from a normal probability distribution with variance specified in a number of ways. This allowed us to study how accurately each regression framework estimated savings in the face of small and large variation as well as autocorrelation and heteroscedasticty. It also allowed us to generate numerous simulated datasets for each facility type to test how the framework performs on average, given random variation in the data.

In *Figure 1*, we provide a visualization of the relationships between energy drivers and energy consumption in the baseline period for the simple facility.
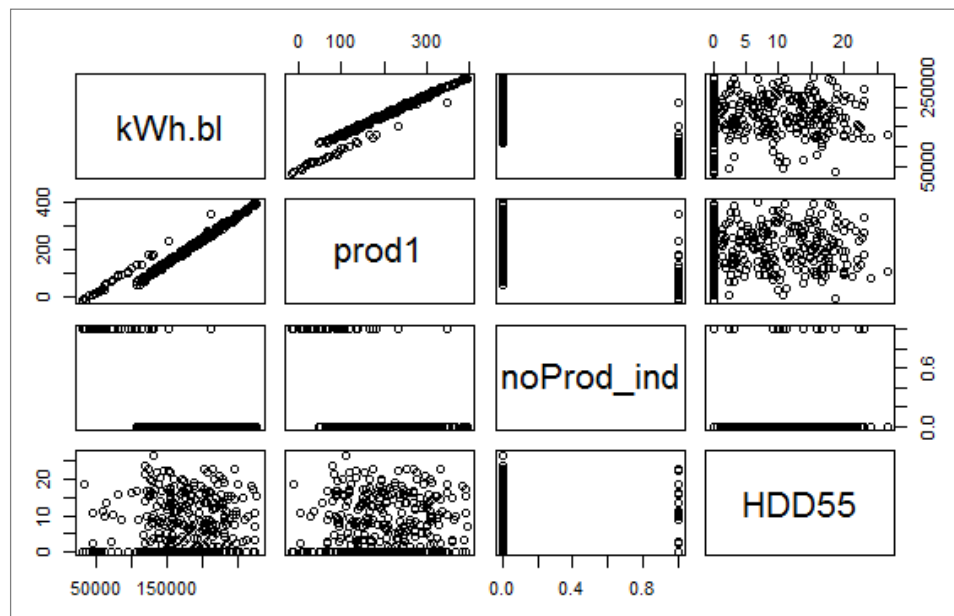

Figure 1: Scatterplot matrix of variable correlations for the simple facility.

Energy consumption at the complex facility is similar but is additionally correlated with a second production variable, a pre-period event indicator, and an interaction between HDD and production. In Figure 2, we present a time series plot showing energy consumption at the complex facility for 365 days before and 365 days after program implementation.
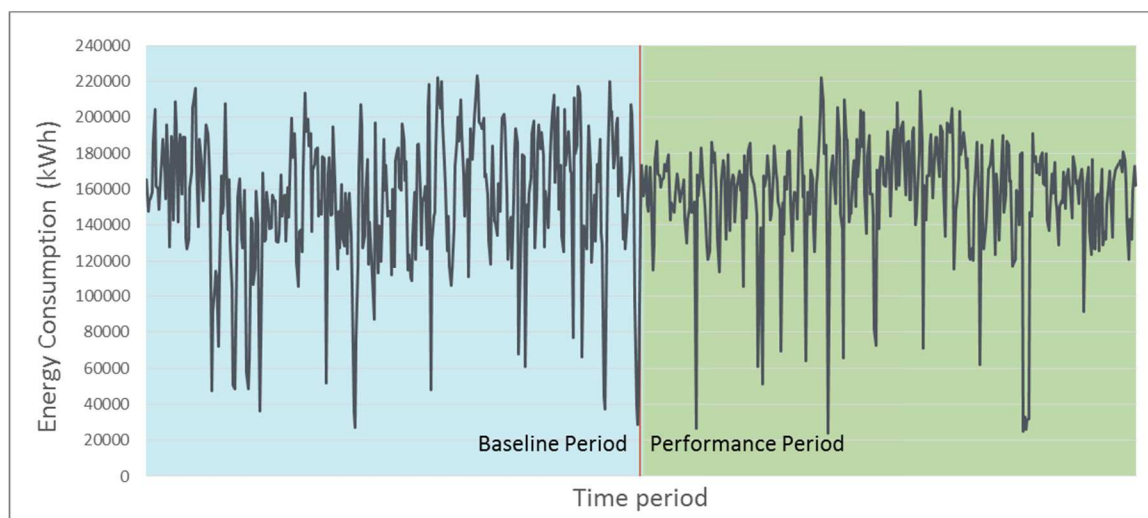

Figure 2. Time series plot of energy consumption during the baseline and post-program periods.

Testing Models

After simulating the facility data, we applied the regression frameworks outlined above to determine which produced accurate results and which performed better for each facility. We considered scenarios where data weren't available on one or more of the production variables, data were missing for

the non-routine adjustment, data were available for an extraneous variable that didn't drive energy consumption, and models where the error terms accounted for or did not account for heteroscedasticity and serial correlation. In summary, we tested each regression framework using regression model specifications with the following characteristics, or cases:

1. Correctly specified
2. Missing event data
3. Omitted weather variable(s)
4. Omitted production2 variable (complex facility only)
5. Extraneous variable (CDD)
6. Heteroscedastic error
7. Serial correlation in errors
8. Known post-program period event with no engineering estimate (i.e., changes to facility consumption due to the event cannot be separated from changes in consumption due to the program in the forecast framework)

In Table 1, we summarize each case by indicating which predictor variables are included in the evaluation regression model; the black dots indicate that the variable was included in the regression model. Some variables (and case four) do not apply in the simple facility; those cells are greyed out in the table.

Table 1. Energy drivers used for each test scenario

Case Number	Energy Drivers Included in Evaluation Regression Model								
								Event in post-period	
	Production 1	Production 2	Interruptions to Production	HDD	Extraneous Variable (CDD)	HDD x Production 1	Event in Baseline	Event Savings Indicated (pre/post)	Event Savings Subtracted (forecast)
Simple Facility									
1	●	(grey)	●	●		(grey)	●	●	●
2	●	(grey)	●	●		(grey)	●		
3	●	(grey)	●			(grey)	●	●	●
4	(grey)	(grey)	(grey)	(grey)	(grey)	(grey)	(grey)	(grey)	(grey)
5	●	(grey)	●	●		(grey)	●	●	●
6	●	(grey)	●	●	●	(grey)	●	●	●
7	●	(grey)	●	●		(grey)	●	●	●
8	●	(grey)	●	●		(grey)	●	●	
Complex Facility									
1	●	●	●	●		●	●	●	●
2	●	●	●	●		●	●		
3	●	●	●				●	●	●
4	●		●	●		●	●	●	●
5	●	●	●	●	●	●	●	●	●
6	●	●	●	●		●	●	●	●
7	●	●	●	●		●	●	●	●

Case Number	Energy Drivers Included in Evaluation Regression Model								
								Event in post-period	
	Production 1	Production 2	Interruptions to Production	HDD	Extraneous Variable (CDD)	HDD x Production 1	Event in Baseline	Event Savings Indicated (pre/post)	Event Savings Subtracted (forecast)
8	●	●	●	●		●	●	●	

We used these 16 total cases to estimate savings in each regression framework, resulting in 48 sets of results. Each set of results included savings estimation and measurements for testing model accuracy for 10,000 simulated data sets we generated for each facility type and error specification. We summarize the findings below.

Findings

We fit regression models according to each of the 16 cases described above and in all three regression frameworks to produce 48 sets of results for each of the 10,000 simulated data sets. The results corresponding to each data set included an estimate of energy savings and an 80% confidence interval. We compared the confidence interval to the "true" savings value to determine whether it included the true savings value and then tallied the number of data sets for which this was true to estimate a coverage rate. For example, if the confidence interval around the savings estimate included the true savings in 9,000 of the 10,000 data sets, we would conclude the method has a 90% coverage rate. Figure 3 illustrates this concept using 100 savings estimates and their associated confidence intervals for the correctly specified complex facility forecast regression framework. In this figure the horizontal axis represents the true savings, each dot represents one savings estimate, and the lines extending from the dots represent the 80% confidence interval for the respective savings estimate.
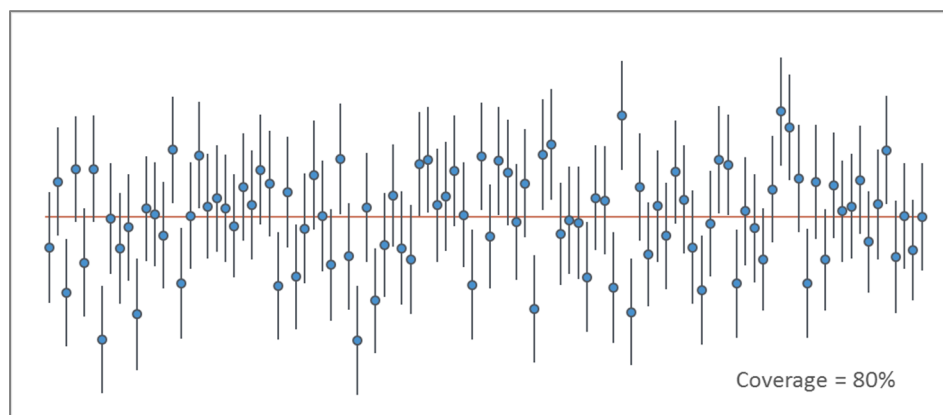

Figure 3. Coverage plot for 100 observed savings estimates from the complex facility using the forecast regression framework.

With 80% confidence intervals, we expect coverage rates to be close to 80% but examine how high or low they are under the various modeling specifications or mis-specifications defined by the cases. Coverage is an important indication of how accurate the results are from each framework – the closer to 80% coverage the framework results in, the more accurate its estimates are.

We summarize our findings in Table 2 where we provide the true savings value, the average estimated savings from each framework, and the 80% confidence interval coverage.

Table 2. Estimated Savings and Confidence Interval Coverage

Case Number	True Savings (MWh)	Estimated Savings (MWh)			80% Confidence Interval Coverage		
		Forecast	Simple Pre/Post	Fully-specified Pre/Post	Forecast	Simple Pre/Post	Fully-specified Pre/Post
Simple Facility							
1	4,508	4,507	4,539	4,506	80%	80%	80%
2	4,508	4,779	4,737	4,780	24%	35%	24%
3	4,508	4,416	4,370	4,337	74%	70%	63%
4	N/A	N/A	N/A	N/A	N/A	N/A	N/A
5	4,508	4,506	4,537	4,505	79%	79%	80%
6	4,508	4,506	4,537	4,504	79%	79%	79%
7	4,508	4,510	4,540	4,507	41%	42%	41%
8	4,508	4,777	4,542	4,510	25%	80%	80%
Complex Facility							
1	5,988	5,987	5,091	5,986	80%	1%	80%
2	5,988	6,256	4,738	6,256	38%	0%	38%
3	5,988	6,222	6,714	6,559	45%	19%	38%
4	5,988	6,378	5,105	5,878	27%	3%	92%
5	5,988	5,987	5,060	5,987	80%	1%	79%
6	5,988	5,986	5,093	5,988	78%	8%	80%
7	5,988	5,983	5,088	5,983	41%	20%	42%
8	5,988	6,254	5,091	5,986	38%	1%	80%

From these results we see that the fully specified pre/post framework provides coverage fairly close to 80% in more cases than the other two frameworks. None of the frameworks have good coverage in Cases 2, 3, or 7 for either of the facilities and the forecast and simple pre/post models do not have good coverage for case 4 in the complex facility. Case 2 represents the scenario where event data are missing and not included in the regression model. Cases 3 and 7 represent scenarios where weather data are missing or omitted from the regression model and where serial correlation exists. Case 4 represents the scenario (in the complex facility only) where one of the production variables is missing or omitted from the model. This implies that none of the frameworks examined in this study can produce accurate savings estimates reliably when variables are omitted or when serial correlation is not accounted for.

We also investigated the bias in savings estimation by computing the mean absolute percentage error (typically referred to as MAPE) and the median percentage error. The MAPE tells us the average magnitude of the estimation bias. The median percentage error tells us whether the model tends to over-predict or under-predict savings. We summarize these results in Table 3.

Table 3. Mean Absolute Percentage Error and Median Percentage Error

Case Number	MAPE			Median Percentage Error		
	Forecast	Simple Pre/Post	Fully-specified Pre/Post	Forecast	Simple Pre/Post	Fully-specified Pre/Post
Simple Facility						
1	2.4%	3.0%	2.9%	0.0%	0.7%	-0.1%
2	6.1%	5.2%	6.1%	6.0%	5.1%	6.0%
3	2.9%	3.9%	4.4%	-2.0%	-3.0%	-3.7%
4	N/A	N/A	N/A	N/A	N/A	N/A
5	2.5%	3.0%	3.0%	-0.1%	0.6%	-0.1%
6	3.5%	4.3%	4.2%	-0.1%	0.6%	-0.1%
7	7.0%	8.4%	8.4%	0.1%	0.7%	0.0%
8	6.0%	3.0%	2.9%	6.0%	0.7%	0.0%
Complex Facility						
1	2.2%	15.0%	2.5%	0.0%	-15.0%	0.0%
2	4.6%	20.9%	4.6%	4.5%	-20.9%	4.5%
3	4.1%	12.1%	9.5%	3.9%	12.1%	9.5%
4	6.5%	14.7%	2.9%	6.5%	-14.7%	-1.8%
5	2.3%	15.5%	2.6%	0.0%	-15.5%	0.0%
6	3.3%	14.9%	3.7%	0.0%	-15.0%	0.0%
7	6.3%	15.3%	7.1%	0.0%	-15.0%	-0.1%
8	4.6%	15.0%	2.5%	4.4%	-15.0%	0.0%

From these results we see that, on average, the absolute percentage error is typically within 5% of the true savings in most scenarios. Similar to the confidence interval coverage results, we see that when variables are missing or omitted and when autocorrelation is not accounted for, the absolute percentage error tends to be elevated. The bias in estimation does not consistently over or under-predict, rather the direction of the bias depends on which variable is omitted. Missing or omitted variables led to either over or under-estimation in both the simple and complex facilities for the forecast and fully specified pre/post regression frameworks, with one exception. When the second production variable was omitted from the complex facility model specification, the fully specified pre/post framework produced unbiased estimates while estimates from the simple pre/post and forecast frameworks were biased. *Figure 4* provides the Mean Absolute Percent Errors (MAPE) for the simple and complex facilities, respectively, and compares values between the three methods of savings estimation.
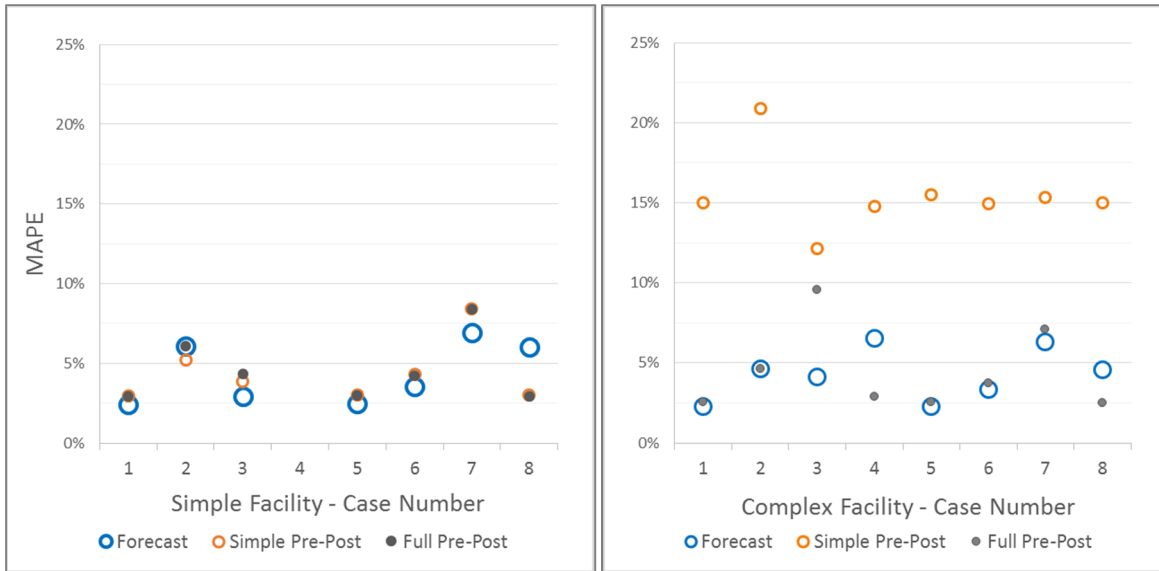
Figure 4. Mean absolute percentage errors for each facility and regression framework.

Consistent with the capture rates, the MAPE value around savings estimated by simple pre/post method are consistently larger for the complex facility. Additionally, in the complex facility the median percentage errors showed a tendency to underestimate savings in all but the omitted weather variable scenarios for the simple pre/post framework. This provides further evidence that the simple pre/post regression framework produces unreliable and biased savings estimates.

Two of the scenarios were created to violate the typical regression assumption that model error is distributed normally. The first case included a heteroscedastic error, where variance increases with production. The second case included serial correlation in the model error as autoregressive order 2. Figure 5 depicts a visualization of the heteroscedastic model error in the complex facility and Figure 6 depicts model error for the first 365 days in the complex facility data.
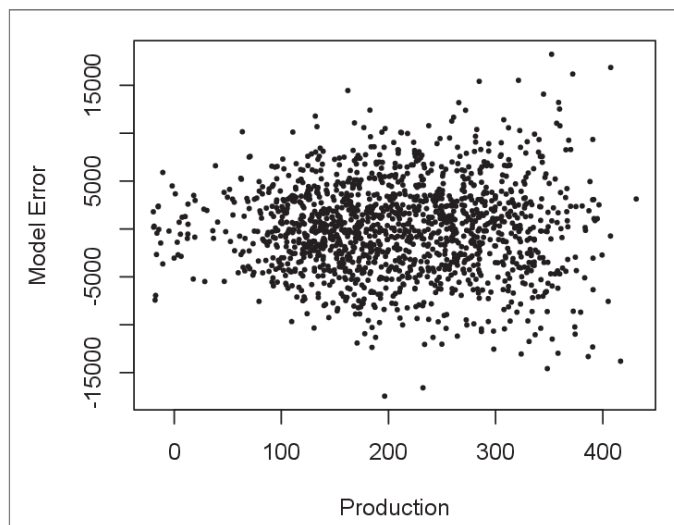

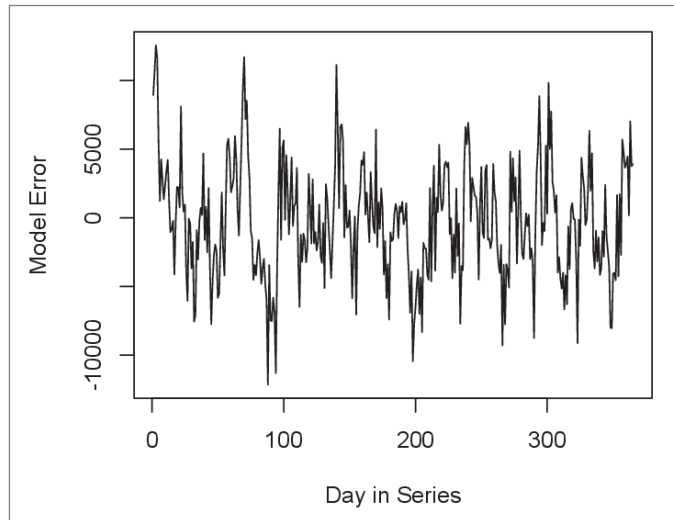Figure 5. Heteroscedastic model error plotted versus production.

Figure 6. Serial correlation in the model errors for the first 365 days in the complex facility.

Despite violation of the normality assumption, the three regression frameworks produced unbiased estimates of energy savings. In both scenarios, the standard errors were larger than models with normally distributed errors. In Table 4, we present the coefficient of variation (CV) for each of the energy savings estimates. The CV is calculated as the margin of error for an estimate divided by the estimate itself. A larger CV implies that the error for that particular estimate is increased.

Table 4. Coefficient of Variation for Simple and Complex Facilities

Case Number	CV - Simple Facility			CV - Complex Facility		
	Forecast	Simple Pre/Post	Fully-specified Pre/Post	Forecast	Simple Pre/Post	Fully-specified Pre/Post
1	3.0%	3.7%	3.7%	2.8%	6.7%	3.2%
2	2.9%	2.9%	2.9%	2.7%	6.4%	2.7%
3	3.3%	4.1%	4.1%	2.7%	6.5%	6.2%
4	N/A	N/A	N/A	3.5%	8.2%	5.2%
5	3.0%	3.7%	3.7%	2.8%	6.8%	3.2%
6	4.3%	5.3%	5.3%	4.1%	7.8%	4.6%
7	3.7%	4.6%	4.5%	3.4%	7.3%	3.9%
8	2.9%	3.7%	3.7%	2.7%	6.7%	3.2%

Conclusions

There are several important conclusions to be drawn from the results of this simulation.

Evaluators should ensure they are including all key drivers of energy in their regression models. Regardless of which framework we use to estimate savings, we see that omitting significant energy drivers from the regression models (i.e., post-program period events, production, and weather) leads to biased estimates of savings, increased standard errors, and 80% confidence intervals that do not actually

include the true savings value 80% of the time. Additionally, there were no apparent consequences for inclusion of an extraneous parameter, so that overspecification of the energy model may not be a concern. The primary take-away from these results is that evaluators should put significant effort into attempting to identify all key energy drivers at a facility and not be overly concerned with inclusion of extraneous variables in the model.

The fully specified pre/post models result in accurate, unbiased estimates under most scenarios. The fully specified pre/post framework produced unbiased estimates for six of the eight scenarios in the complex facility data and for five of the seven scenarios in the simple facility data, outperforming the other regression frameworks. It was also the only framework to produce an unbiased result in one of the three variable omission scenarios for the complex facility. For this reason we conclude that of the frameworks tested, the fully specified pre/post was the most robust model for savings estimation.

The simple pre/post model is unreliable and should not be used to estimate facility energy savings. The simple pre/post model performed consistent with the forecast and fully specified pre/post model for the simple facility data, however it failed to produce accurate, unbiased estimates nor capture the true savings at the target capture rate with the complex facility data. In practice, industrial facilities tend to be complex, with many potential drivers of energy and interactions. It is unlikely that an actual facility will behave like the simple facility modeled in this study, therefore we do not recommend use of the simple pre/post regression framework.

Evaluators should investigate serial correlation and attempt to model it in their chosen regression framework. Many industrial facilities have production that is seasonal or cyclic. It is likely then, that there is some degree of serial correlation present in the data from these facilities. This is particularly true when the data is higher frequency, such as daily intervals. Our results suggest that failing to account for this autocorrelation can greatly reduce the chances of capturing the true savings parameter in the estimated confidence interval.

Only the fully specified pre/post regression framework should be used when an event or non-routine adjustment occurs in the post-program period and for which no estimate of energy reduction or increase exists. In our experience as evaluators, we have encountered several cases where a facility underwent a significant change during the post-program period. In some cases, such as new equipment installations or upgrades, we will have an engineering estimate available that quantifies the change in energy consumption resulting from the installation. For many of these situations, however, we have no estimate for the expected increase or reduction in energy consumption that results from a change to the facility energy consumption. Some examples we have encountered are: layoffs at the facility, temporary building closures, changes in management, or equipment breakdowns. In this simulated scenario, the forecast framework overestimated savings by including the reduction in consumption due to the event in the total savings attributed to the SEM program. Conversely, the fully specified pre/post framework was able to account for the reduction in consumption by way of an added indicator variable, resulting in unbiased estimates of savings.