A Comparison of Approaches to Estimating the Time-Aggregated Uncertainty of Savings Estimated from Meter Data

Bill Koran, SBW Consulting, West Linn, OR
Erik Boyer, Bonneville Power Administration, Spokane, WA
M. Sami Khawaja, Ph.D., The Cadmus Group, Portland, OR
Josh Rushton, Ph.D., Rushton Analytics, Portland, OR
Jim Stewart, Ph.D., The Cadmus Group, Portland, OR

ABSTRACT

An increasing number of utility programs use savings estimates based on changes at the utility meter, using regressions or other data-driven approaches. The program types include pay-for performance, strategic energy management, some of the behavior-based programs, and programs fitting the new requirements in California using changes in "normalized metered energy consumption" as a measure of energy savings. Therefore, an understanding of the uncertainty in these estimates is increasingly important. For ordinary least squares regression, the calculations for the uncertainty in the estimate for an individual point are easy to find in statistics literature. The aggregate uncertainty for multiple points—e.g., the uncertainty for a full year of savings—is little discussed.

This paper describes the aggregate uncertainty for multiple points using several different methods: the approach described in ASHRAE Guideline 14-2014 ("Measurement of Energy, Demand, and Water Savings"), a revised ASHRAE approach, an exact algebraic approach for ordinary least squares regression, and bootstrap approaches. Savings uncertainties estimated using the different approaches are compared for both synthetic data and energy use data from a real building.

A recommended equation for a simplified approach to estimating uncertainty is provided, and the utility of a bootstrap for estimating uncertainty is shown.

Introduction

Along with the growth in whole building programs by utilities, there seems to be increasing overlap between measurement and verification (M&V) and program impact evaluation. Credible site-specific M&V can ease evaluation burdens. However, the former is often the domain of energy engineers and analysts, who often lack significant statistical expertise. The latter is often the domain of statisticians and economists, or other people with statistical expertise, but who lack engineering knowledge.

In the opinion of the authors, it is valuable for engineers and statisticians to learn from each other. Statistical knowledge can help engineers not only quantify energy savings, but also provide early verification of performance of measures, identify when a site's energy use is changing, provide fault detection, and in some cases, even diagnose faults.

Similarly, statisticians may find value in understanding how a measure might save energy, and how those processes might show up in an energy model. To maximize the benefits of regression and other data-driven models, the authors believe such models should usually have physical significance.

Background

As described in the International Performance Measurement and Verification Protocol (IPMVP), avoided energy consumption or demand quantifies savings in a reporting period relative to the energy that would have been consumed without the efficiency measure(s). To estimate savings, a model for the pre period is built using energy use as the dependent 'y'-variable and some independent 'x'-variable (e.g., weather) expected to explain most of the variation in use as the. Once the model is built, it can be used to forecast, or predict, energy use in the post or reporting period. This is the estimate of "what energy use would have been had the energy saving measure not been installed." The difference between this prediction and the actual use is the estimated energy savings or "avoided energy use."

The adjusted baseline energy use is the sum of the forecasts for each of the points (measurement periods) in the reporting period. For example, if monthly billing data are used, it is the sum of the forecasted energy use for each monthly billing period within the reporting period. These forecasts are typically made with a statistical regression model. The reported savings are the adjusted baseline minus the actual metered energy.

To evaluate the uncertainty in any value derived from regression fit, one must first determine whether the value should be thought of as an *estimate* or a *prediction*. In the present context, the distinction is this:

1. *Estimation* is concerned with the *average,* or *expected,* y-value, given a specific x-value.
2. *Prediction* is concerned with the *specific* y-value that may accompany a specific x-value.

A confidence interval quantifies the uncertainty in an estimate, for an input confidence level. For the case where an ordinary least squares (OLS) regression fit is used to estimate the *average,* or *expected*, y-value associated with a particular x-value, the formula for the confidence interval (CI) can be found in any regression textbook:

$$\mathrm{CI} = \hat{y} \pm t \times s \times \sqrt{1/n + (x_0 - \bar{x})^2 / (\sum (x_i - \bar{x})^2)}$$

Where:

$\hat{y}$ is the point estimate for the *expected* y-value, given x_0
t is the Student's t-statistic for the input confidence level and number of points in the data
s is the standard error of the noise terms, estimated by the Root Mean Squared Error (RMSE)
n is the number of points (x-y pairs) used to fit the regression model
x_0 is the particular value of the independent variable for which the mean y-value is estimated
x_i are the individual independent variable values that were used to fit the regression
$\bar{x}$ is the mean of the x_i

A prediction interval (PI) quantifies the uncertainty in the prediction of an individual point—the uncertainty in predicting an individual y-value for a given x-value, at an input confidence level:

$$\mathrm{PI} = \hat{y} \pm t \times s \times \sqrt{1 + 1/n + (x_0 - \bar{x})^2 / (\sum (x_i - \bar{x})^2)}$$

Prediction Intervals are much wider than Confidence Intervals, since it is much harder to predict the value of an individual point than it is to predict the typical or average value of a point under the same conditions.

The *uncertainty* in the adjusted baseline energy use is the combination of the uncertainty in each of the forecasts for the points in the reporting period. The uncertainty in estimated savings, for IPMVP "Avoided Energy Use," is simply the uncertainty of the baseline model forecast of what would have been the energy use during the reporting period, had no changes occurred in the building, the "adjusted baseline." For monthly billing data, if the uncertainty in a year of savings is needed, the uncertainty in the adjusted baseline for January is combined with the uncertainty in the adjusted baseline for February, and for March, etc.

There are at least two complications in estimating the overall forecast and savings uncertainty:

1. Uncertainty in savings estimates requires the combination or aggregation of predictions for each measurement interval. For example, with monthly billing data, there is an uncertainty associated with the baseline prediction for each billing period. What is the combined uncertainty for all of these predictions? The answer to this is less available in the literature.
2. Many of the models used for estimating baseline energy use, and hence savings, violate the assumptions for OLS regression, so the above equations may not be applicable.

Scope

This paper compares the results from several methods of estimating the uncertainty of regression models' savings estimates. Each uncertainty method was tested for several different data sets. For each data set, an appropriate regression model was fit, predictions were made based on that model, and aggregated prediction uncertainty was estimated using multiple methods

Four data sets were analyzed for the uncertainty in aggregated predictions from regression models. These data sets can be summarized as follows:

1. Synthetic data, linear relationship, no autocorrelation
2. Synthetic data, linear relationship, moderate autocorrelation
3. Synthetic data, linear relationship, higher scatter, higher autocorrelation
4. Real data for 24 months, 4-parameter relationship, X-values not independent

The first data set met all of the requirements for ordinary least squares regression, and was used to check all of the methods. The fourth data set used real monthly data with a distinct change in slope. Hence, the least squares fit was for a 4-parameter model with a change point.

The analyses for each data set were performed using multiple methods. Four primary methods were used, with variants for specific data sets.

1. Algebraic solution for aggregated uncertainty from OLS regressions. To handle data sets with autocorrelation, the equations were adjusted using a modification of the Reddy and Claridge (2000) approach.
2. ASHRAE Fractional Savings Uncertainty (FSU), from Guideline 14 and Reddy and Claridge (2000)
3. Improved FSU, from Sun and Baltazar (2013)
4. Bootstrap Resampling
    a. Resample data X-Y pairs
    b. Block Bootstrap for autocorrelated residuals
    c. Resample residuals

Methodology

This section describes the four data sets and the four methods used for estimating the aggregated uncertainty associated with predictions for multiple time intervals from a regression model.

Data Sets

The four data sets and models are shown in Figures 1 through 4 below. The scatter charts show the data, the fit line, and dots indicating the 90% prediction intervals. Autocorrelation plots are shown for those interested in the degree of autocorrelation. Each of the first three data sets has 365 points (X-Y pairs), representing daily data. The fourth data set is monthly data, and hence doesn't have autocorrelation. No axis scales are shown because this is customer data.

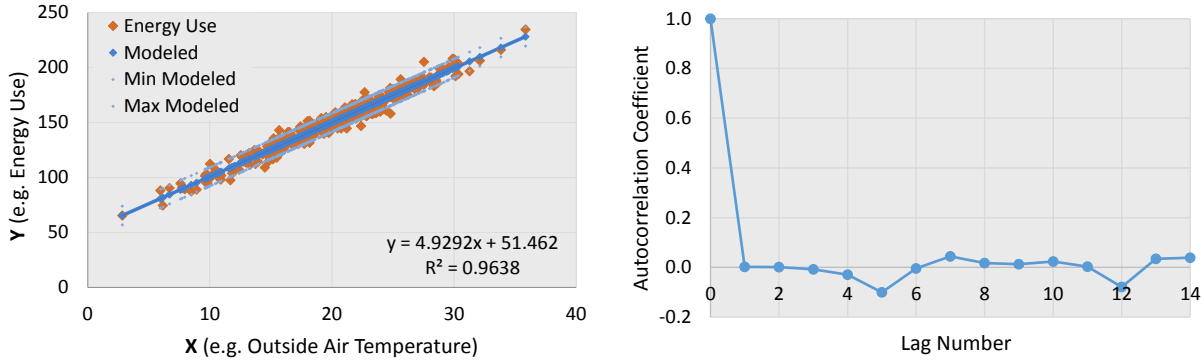

Figure 1. Synthetic data, linear relationship, no autocorrelation

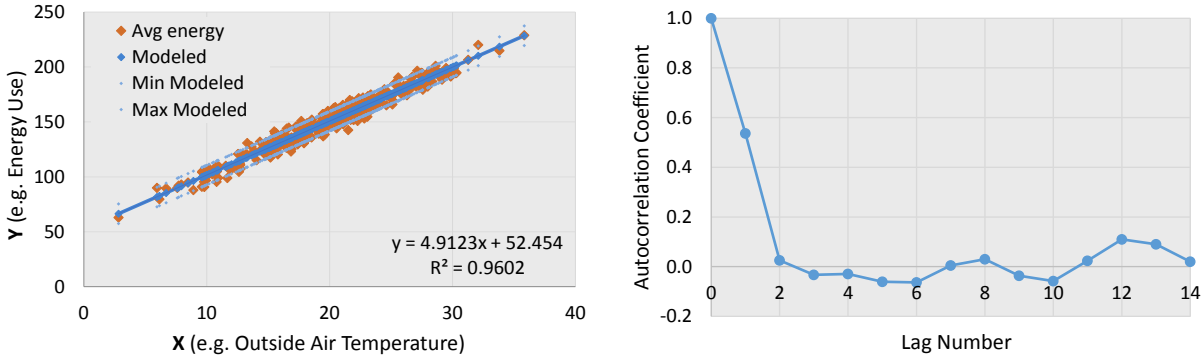

Figure 2. Synthetic data, linear relationship, Lag-1 autocorrelation

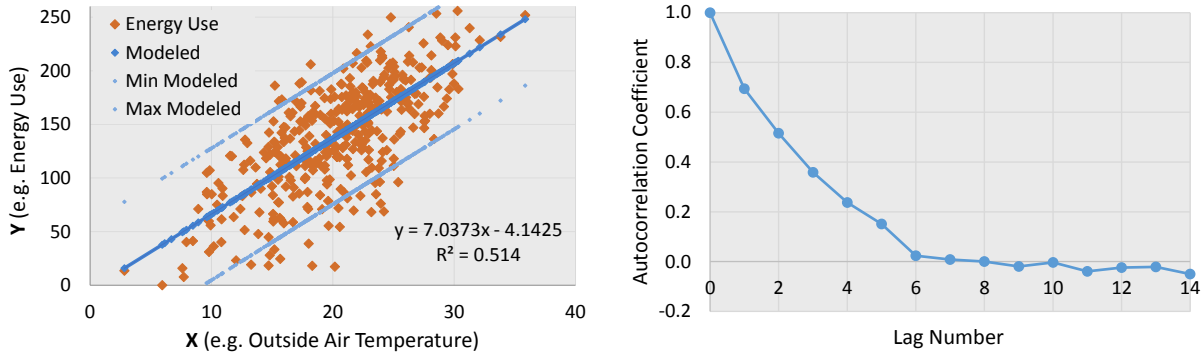

Figure 3. Synthetic data, linear relationship, higher scatter, significant autocorrelation

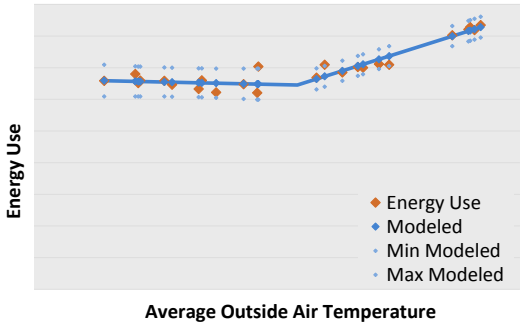

Figure 4. Monthly Data With a 4-Parameter Relationship

ASHRAE Guideline 14 Equation

This is intended for models with weather as the primary independent variable. It includes an adjustment for models with correlated residuals—those where each point has a relationship with the points associated with recent prior timestamps. Correlated residuals (known as time-series autocorrelation) are common when the measurement time interval is short, such as hourly or daily.

$$\frac{\Delta \mathrm{E}_{\mathrm{save,m}}}{\mathrm{E}_{\mathrm{save,m}}} = \mathrm{t} \times \frac{1.26 \times \mathrm{CV} \sqrt{\left[\frac{n_{base}}{n'_{base}}\left(1+\frac{2}{n'_{base}}\right)\frac{1}{m_{post}}\right]}}{\mathrm{F}}$$

Where:

$\frac{\Delta \mathrm{E}_{\mathrm{save,m}}}{\mathrm{E}_{\mathrm{save,m}}}$ is the uncertainty in the savings divided by the savings

CV is the coefficient of variation of the root mean squared error CV(RMSE).

$$CV(RMSE) = \frac{\sqrt{\left[\sum_i (E_i - \widehat{E}_i)^2/(n-p)\right]}}{\overline{E}}$$

F = Savings fraction=$E_{save}/E_{baseline}$

t = Student's t-statistic

E_i = measured energy use for time interval *i*

$\widehat{E}_i$ = predicted energy use for time interval *i*

$\overline{E}$ = average energy use for all time intervals

n_{base} = number of points in the baseline period

m_{post} = number of points in the post period

p=the number of model parameters

$n'_{base} = n_{base} \times \frac{1-\rho}{1+\rho}$ is the effective number of points after accounting for autocorrelation.

ρ is the autocorrelation coefficient, which is the square root of the R^2 calculated for the correlation between the residuals and the residuals for the prior time period.

1.26 is an empirically derived factor to eliminate the need for matrix algebra. The derivation is described in Reddy and Claridge 2000.

Improved Approach Based on Guideline 14

This is an improved formula developed by Sun and Baltazar. They showed that the deviation of the simplified formula's uncertainty estimates from the true value was in the form of a polynomial. Therefore, their formula replaces the 1.26 factor with a second order polynomial, using the number of

months in the reporting period as a variable in the polynomial. There are different coefficients depending upon whether the data have daily or monthly intervals.

The improved formula is:

$$\frac{\Delta E_{save,m}}{E_{save,m}} = t \cdot \frac{(aM^2 + bM + c) \cdot CV\sqrt{\left[\frac{n}{n'}\left(1 + \frac{2}{n'}\right)\frac{1}{m}\right]}}{F}$$

Where:

M=number of months of reporting period data

a, b, and c were solved for using the deviation of the Guideline 14 uncertainty estimates from the true value estimated by matrix algebra for OLS regressions. Their values are:

Data Interval	Monthly	Daily
a	-0.00022	-0.00024
b	0.03306	0.03535
c	0.94054	1.00286

Exact Equation for Ordinary Least Squares Regression

An exact equation for the aggregated uncertainty of predictions for many time intervals for OLS regressions can be derived from the standard matrix formula for a linear combination of OLS regression coefficients. To use the formula, the sum of the post-period predicted values is written as a matrix product involving the fitted regression coefficients:

$$\hat{y} = \sum\left(\hat{\beta}_0 + \hat{\beta}_1 x_i^{\text{post}}\right) = \left[m_{post}, \quad \sum x_i^{\text{post}}\right]\begin{bmatrix}\hat{\beta}_0 \\ \hat{\beta}_1\end{bmatrix} = \boldsymbol{x}^{\text{post sum}\prime}\begin{bmatrix}\hat{\beta}_0 \\ \hat{\beta}_1\end{bmatrix} = \boldsymbol{x}^{\text{post sum}\prime}\widehat{\boldsymbol{\beta}}$$

The formula says that the variance of such a product can be expressed as another product involving the regression parameters' covariance matrix:

$$\text{Var}(\hat{y}) = \boldsymbol{x}^{\text{post sum}\prime}\text{Cov}\left(\widehat{\boldsymbol{\beta}}\right)\boldsymbol{x}^{\text{post sum}}$$

The covariance matrix itself is a standard result that can be found in most regression textbooks. Using these results, we can derive explicit formulas for (1) the standard error of $\hat{y}$ (i.e., the component of prediction uncertainty that is due to uncertainty in the model fit itself), and (2) the component of prediction uncertainty that is due to typical noise in the post period.

Space limitations preclude including the full derivation[1] here. The derivation shows that there are two components to the overall uncertainty in the adjusted baseline energy use. They are:

Aggregated Uncertainty in the Model Fit. This is the uncertainty in the mean total consumption for the reporting or post period):

$$t \times se(\hat{y}) = t \times \frac{s \times m_{post}}{\sqrt{n_{base}}} \times \sqrt{\left(1 + \frac{\left(\bar{x}_{base} - \bar{x}_{post}\right)^2}{\text{Var}(x_{base})}\right)}$$

Where:

$se(\hat{y})$ is the standard error aggregated for all the points in the reporting period

s is the standard error of the model's noise terms, estimated by the regression RMSE

[1] Josh Rushton, Ph.D., derived the expression. His derivation will be available in the forthcoming Koran 2017.

$\bar{x}_{base}$ is the mean of the *x* -values in the baseline period
$\bar{x}_{post}$ is the mean of the *x* -values in the post period
$\mathrm{Var}(x_{base})$ is the variance of the *x*-values in the baseline period

Aggregated Uncertainty due to Noise. This is simply the baseline standard error aggregated by quadrature for the number of points in the post period:

$$t \times \sqrt{s^2 \times m_{post}} \;=\; t \times s \times \sqrt{m_{post}}$$

Total Adjusted Baseline Uncertainty. This is the combination of the model and noise uncertainties:

$$t \times se\left(\hat{y} + \sum \varepsilon_i\right) = t \times s \times \sqrt{\frac{{m_{post}}^2}{n_{base}} \times \left(1 + \frac{\left(\bar{x}_{base} - \bar{x}_{post}\right)^2}{Var(x_{base})}\right) + m_{post}}$$

In the presence of autocorrelation, exact closed-form solutions are not generally available even for point estimates so statisticians typically employ numerical approximation methods. To approximate the impact of autocorrelation, the authors propose adjusting m_{post} by the ratio of n_{base}/n'_{base} similar to the ASHRAE FSU approach. Empirical checks for the data sets in this work suggest that this adjustment should only be applied to the noise portion, and the final equation is:

Total Adjusted Baseline Uncertainty with Approximate Autocorrelation Adjustment. M&V practitioners needing a simplified method to estimate uncertainty should use this equation.

$$t \times s\left(\hat{y} + \sum \varepsilon_i\right) = t \times s \times \sqrt{\frac{{m_{post}}^2}{n_{base}} \times \left(1 + \frac{\left(\bar{x}_{base} - \bar{x}_{post}\right)^2}{Var(x_{base})}\right) + m_{post} \times \frac{n_{base}}{n'_{base}}}$$

Note that model and noise uncertainty aggregate differently. The model uncertainty aggregates directly with m_{post} whereas the noise uncertainty aggregates with the square root of m_{post}. This explains why the ASHRAE approach was improved by changing the fixed coefficient of 1.26 with a polynomial.

Ruch, Kissock, and Reddy noted that "A common but incorrect practice for computing prediction error bounds of the sum of daily predictions is to sum in quadrature the prediction error bounds of the daily predictions… However…all of the daily predictions of energy use are based on the same estimated regression coefficients and are therefore correlated. Consequently, summing in quadrature will underestimate the correct prediction error bound."

In layperson's terms, to the extent the model is different from the true relationship, it is different for every prediction in the same way, and hence the model uncertainty aggregates with m_{post}. In contrast, the noise uncertainty is random, so it aggregates with the *square root* of m_{post}.

Bootstrap Approaches

The bootstrap approaches are described in greater detail than the other methods, since they do not seem to be in common use for estimating the uncertainty in energy savings from regressions. We tend to accept the bootstrap's uncertainty estimates as relatively sound because of the method's directness and transparency. In other words, agreement with the bootstrap, or block bootstrap for data sets with autocorrelation, is taken as support for a simpler method, and disagreement is taken as cause for concern. Further study is needed to determine the reliability of such comparisons.

Basic Bootstrap Approach—Resample Data X-Y Pairs. Statistical bootstrapping falls under the broader heading of resampling and it involves a relatively simple procedure repeated many times. Bootstrapping provides a method for estimating confidence intervals when traditional uncertainty equations fail. The underlying assumption for the bootstrap is that the distribution of a parameter (e.g. model slope, or savings estimate) in the sample closely approximates its distribution in the population, and that the distribution in the bootstrap *sample of the sample* (resample) closely approximates its distribution in the *original sample*, i.e. the bootstrap sample is a close representation of the sample, and the sample is a close approximation of the population.

In a bootstrap, a sample is drawn, with replacement, from the original sample. Parameters are estimated; savings are estimated; the bootstrap sample is placed back into the original sample 'deck,' and the process is repeated many times. Therefore, for each bootstrap sample, some items can be drawn multiple times, whereas others may not be drawn at all, and each bootstrap sample has the same *number* of members as the original sample.

After getting the distributions for the coefficients and each component of uncertainty, the 5th percentile values were subtracted from the 95th percentile values used to get the ranges of coefficients and errors for a 90% confidence level.

Block Bootstrap Approach for Autocorrelated Residuals. A regular bootstrap does not retain the order of the data. For example, starting with a data set of 365 consecutive days, a bootstrap sample of 365 days could have a starting sequence of March 7, January 31, September 10, and March 20. Hence, any time series relationships are lost. If residuals are autocorrelated, then the time series nature of the data is important and must be preserved. One solution to this is to resample the data in blocks. For example, with daily data, two or more days will be treated as a block and kept together when resampling.

These analyses were performed for data sets 2 and 3.

A key question with autocorrelation and a block bootstrap is, how big a block to use? For how many days is the energy use related? Is it only related to the prior day, i.e., lag 1 for daily data? Is it related to the prior 4 days? Block sizes of 1, 2, 4, 7, and 14 points (e.g. days) were tested to determine when the estimated aggregated error ceased to increase with increasing block size.

Resample Residuals—A Bootstrap Approach for Data with a Relationship Between Independent Variable Values. If the independent variable is outside air temperature, there is a certain distribution of these temperatures that will occur over the course of a year. The X-values are related because of this distribution. Ignoring this constraint while bootstrapping could cause the estimated uncertainty to deviate from the true uncertainty.

A solution is to resample residuals instead of data pairs: Add the residuals to the fitted lines of a model built from the original data to create a new data sample, create new models, and proceed as normal. To look at the impact of the X-values having a relationship with each other, three bootstrap analyses were performed:

1. Resample data pairs as described in the prior section.
2. Resample the residuals based on the fit of a least squares regression for the original data set.
3. Resample a normal distribution of residuals based on the fit and standard error of a least squares regression for the original data set.

Approaches 2 and 3 retain the distribution of the temperatures. With Approach 1, a resampling of the data pairs, a bootstrap sample could have too many hot temperatures and cold temperatures, and too few mid temperatures, and hence give an unreasonable estimate of savings and uncertainty.

The implementation for Approach 3 used the RMSE for the total model as the center of the normal residuals, and this was multiplied by the t-statistic to get the 90% confidence interval for the normal residuals. The two approaches for bootstrapping residuals were both included because the model residuals for data set 4 exhibited heteroscedasticity, and hence these two approaches were expected to give slightly different results.

Results for Each Uncertainty Estimation Method

Results for Synthetic Data with a Linear Relationship and No Autocorrelation

The uncertainties were estimated for reporting periods from 30 to 365 data intervals. Figure 5 summarizes the results from the estimation of total errors for all four methods.

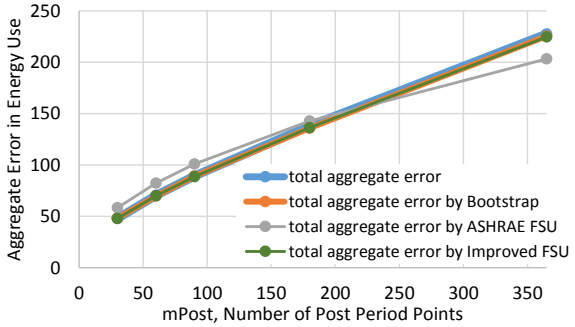

Figure 5. Total Aggregate Errors by Approach and Length of Reporting Period

Note that all of the methods provide essentially the same uncertainty, except the original ASHRAE FSU approach, which slightly overstates the uncertainty for reporting periods of less than about 6 months, and understates it for reporting periods longer than about 8 months.

Results for Synthetic Data with a Linear Relationship and Moderate Autocorrelation

The bootstrap was run for all the block sizes listed above. The results in Figure 6 show that a block size of 4 is needed before the error ceases to increase. As expected, not accounting for autocorrelation underestimates the uncertainty, as shown by the values for OLS.

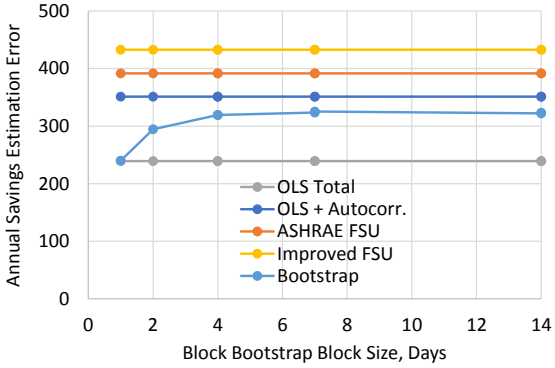

Figure 6. Total Error vs. Block Size for Data Set 2

Also as expected, the ASHRAE FSU equation estimates a lower total uncertainty for a full year of data than the improved equation. This is the same result shown for Data Set 1, and the issue found by Sun and Baltazar in their research. The OLS + Autocorrelation estimate uses a similar adjustment for effective number of points as the FSU equations. However, the equation for OLS + Autocorrelation applies that adjustment only to the noise portion of the equation, so it estimates lower uncertainty than the improved FSU equation, and closer to the result for the block bootstrap. The improved FSU equation estimates an error that is about 33% higher (433/326 – 1) than the bootstrap.

Results for Synthetic Data with a Linear Relationship, Higher Scatter and Higher Autocorrelation

Figure 7 shows the results of the bootstrap runs with blocks of 1, 2, 4, 5, 7, and 14 points.

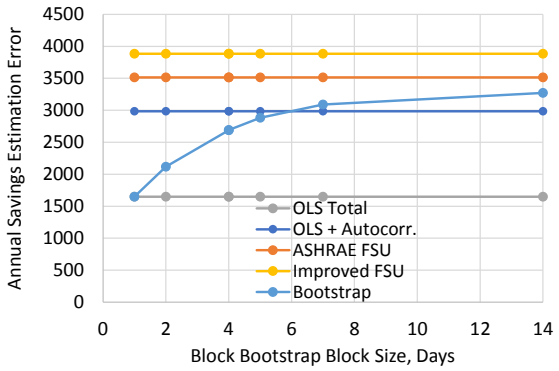

Figure 7. Total Error vs. Block Size for Data Set 3

Here, a failure to account for the increased autocorrelation would vastly underestimate the uncertainty. Using the block size of 7 as the minimum appropriate, and assuming the bootstrap is providing a good estimate of uncertainty, not accounting for autocorrelation would underestimate the true uncertainty by nearly 50%. The OLS + Autocorrelation equation provided results close to the bootstrap, and the FSU equations continue to estimate higher uncertainties.

Results for Real Data with a 4-Parameter Relationship

Figure 8 shows the results for the OLS approach, both FSU methods, and three resampling methods. These analyses were for a 24-month baseline and a 24-month reporting period.

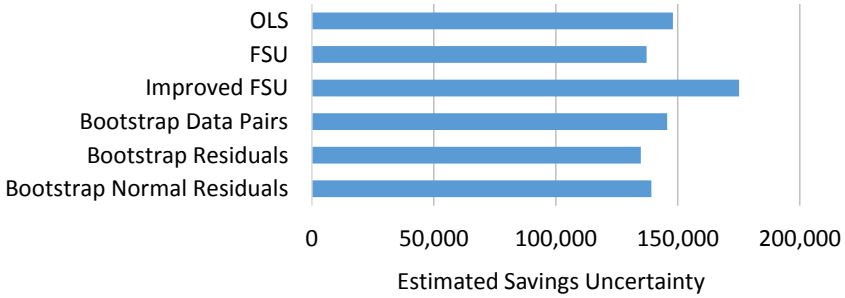

Figure 8. Estimated Uncertainties by Method for a Data Set With Related X-Values

Comparison of the OLS uncertainty estimate with the bootstrap estimates indicates that OLS provides a reasonable estimate for this data set. The OLS approach might not be considered applicable, since the model is not linear. However, its utility for these applications is of interest because the estimation is simple. The OLS estimate of uncertainty was slightly higher than the bootstrap estimates, which were expected to be more reliable. This is not surprising, since OLS assumes that all data pairs are independent, but here the X-values are not independent. However, the OLS estimate was within 10% of the bootstrap estimates, so it may provide a reasonable approximation for some models that are more complex than OLS.

Of the bootstrap methods, the resampling of data pairs had the highest uncertainty. This was expected, since bootstrap samples of data pairs could have mostly cold temperatures, or mostly high temperatures, with attendant higher uncertainties.

The bootstrapping of normal residuals results in an estimated uncertainty that is very close to the estimate for bootstrapping the raw residuals.

The improved FSU shows much higher uncertainty than the OLS approach. The original FSU provides uncertainty estimates close to the bootstrap estimates since, relative to the improved FSU, it underestimates uncertainty for reporting periods longer than 8 months. Since the close estimate to the bootstrap is a result of this error, little confidence should be placed in this approach for other data sets.

Summary, Conclusions, and Recommended Further Work

Summary

The uncertainty for each prediction in a time interval aggregates over the multiple time intervals in a reporting period. For example, with monthly billing data, the annual energy savings estimate is based on 12 predictions. The estimated annual energy savings uncertainty is the combination of the uncertainty for each of these predictions.

Four data sets were analyzed for the uncertainty in aggregated predictions from regression models. Such predictions allow the estimation of uncertainty in cumulative energy savings over a reporting period of multiple metering periods. These data sets can be summarized as follows: (1) Synthetic data, linear relationship, no autocorrelation; (2) Synthetic data, linear relationship, moderate autocorrelation; (3) Synthetic data, linear relationship, higher scatter, higher autocorrelation; and (4) Real data, 4-parameter relationship, X-values not independent.

The analyses were performed using multiple methods. Four primary methods were used, with variants for specific data sets: (1) Algebraic solution for aggregated uncertainty from OLS regressions. To handle data sets with autocorrelation, the equations were modified using the ASHRAE FSU Approach to estimate an effective number of data points; (2) ASHRAE FSU equations from ASHRAE Guideline 14; (3) Improved FSU method, from Yifu Sun and Juan-Carlos Baltazar, Ph.D.; and (4) Bootstrap Resampling.

For Data Set 4, three Bootstrap approaches were used: (a) Resample Data X-Y Pairs; (b) Resample Residuals; and (c) Resample Normal Residuals.

Conclusions

- The derived equation for OLS demonstrates that there are two components of uncertainty to be considered in the aggregation over the reporting period: The model or mean response uncertainty, and the noise (t-statistic times the standard error) uncertainty.
- The improved ASHRAE FSU equation by Sun and Baltazar provides results that match the derived equation for OLS models, but for more complex models can provide very different results.

- A bootstrap approach can be an effective way of estimating uncertainty for complex model formulations, including for models with autocorrelation. When bootstrapping using weather variables, it is better to resample residuals instead of x-y data pairs, but this appears to be a minor consideration.
- Based on the results using the block bootstrap, the ASHRAE approach to using an "effective" *n* to account for autocorrelation seems to overstate the impact of autocorrelation.
- For models with more coefficients than a simple OLS model, if those coefficients are related to each other, the OLS equation can estimate an uncertainty that is too high.

The OLS uncertainty equation, Improved FSU, and Bootstrapping X-Y pairs provide almost identical results for a linear data set without autocorrelation, regardless of the length of the reporting period. The original ASHRAE FSU method deviates for reporting periods under or over about 7 months.

For a data set with autocorrelation, it is well-known that not accounting for that autocorrelation can greatly underestimate the uncertainty. Based on the results for the block bootstrap, it appears that the ASHRAE adjustment to handle autocorrelation may significantly overstate its impact.

For more complex models, the OLS approach may still provide reasonable estimates of uncertainty. For data where the X-values are related, such as energy models based on outside air temperature, the OLS approach appears to slightly overestimate the uncertainty. Resampling residuals or normal residuals gave what are believed to be the best uncertainty estimates for data set 4, but the OLS approach provided an estimate that was less than 10% high.

In the big picture, all of the approaches provided reasonable results. None of them were off from the others by an order of magnitude, or even a factor of two. The equations for estimating uncertainty based on OLS seemed to work fairly well even for data requiring a 4-parameter change point linear model.

Recommended Further Work

The data sets analyzed were relatively simple to model. Three of the data sets could be fit with a linear regression. The fourth data set required a model with 4-parameters.

However, many energy models are more complex. With daily data, it is not unusual to have 2 or 3 daytypes, with each daytype model having 3, 4, or 5 parameters, for a total of 6 to 12 parameters or more in the overall model. With hourly data, models may have even more parameters. It would be beneficial to know how well these methods for estimating uncertainty perform for more complex models and models built on shorter-interval data with greater autocorrelation.

In addition, a more computationally intensive study could directly evaluate the quality of the different uncertainty methods by iteratively simulating baseline and post-period data from known distributions (as with data sets 1, 2, and 3), creating prediction intervals base on baseline data for each method, and checking whether the actual (simulated) post-period values agree with the prediction interval. By repeating this experiment many times, one can estimate the actual coverage rate of the different intervals. A 90% prediction interval should cover the target roughly 90% of the time.

Acknowledgment

The Bonneville Power Administration funded documentation of this work.

References

ASHRAE Guideline 14-2014: Measurement of Energy, Demand and Water Savings, American Society of Heating, Refrigerating, and Air Conditioning Engineers.

International Performance Measurement and Verification Protocol: Core Concepts. 2016. EVO 10000 – 1:2016, Efficiency Valuation Organization.

International Performance Measurement and Verification Protocol: Concepts and Options for Determining Energy and Water Savings Volume I. 2012. Efficiency Valuation Organization.

Koran, W. E. 2017. Uncertainty Approaches and Analyses for Regression Models and ECAM, Bonneville Power Administration.

Reddy, T. A. and Claridge, D. E. 2000. "Uncertainty of 'Measured' Energy Savings from Statistical Baseline Models." HVAC&R Research, January 2000.

Ruch, D.K., Kissock, J.K., and Reddy, T.A., 1999. "Prediction Uncertainty of Linear Building Energy Use Models with Autocorrelated Residuals, ASME Journal of Solar Energy Engineering, vol.121

Sun, Y. and Baltazar, J.C. 2013. "(DE-13-C033) Analysis and Improvement on the Estimation of Building Energy Savings Uncertainty," Proceedings of the 2013 ASHRAE Annual Conference, Denver, CO.