
UNCERTAINTY APPROACHES AND ANALYSES FOR REGRESSION MODELS AND ECAM

Prepared for **BONNEVILLE POWER ADMINISTRATION**
Carrie Cobb

Prepared by **SBW CONSULTING, INC.**
2820 Northup Way, Suite 230
Bellevue, WA 98004

August 11, 2017



ENERGY • WATER • EFFICIENCY

TABLE OF CONTENTS

1. EXECUTIVE SUMMARY	1
1.1. Reason for the Work.....	1
1.2. Summary of the Work.....	1
1.3. Author’s Comment	2
2. BACKGROUND AND LITERATURE REVIEW.....	4
2.1. Foundations of Models and Uncertainty	4
2.1.1. Confidence and Prediction Intervals for Ordinary Least Squares Regressions	4
2.1.1.1. Confidence Level.....	5
2.1.1.2. Confidence Interval	5
2.1.1.3. Prediction Interval.....	7
2.1.1.4. Putting it All Together	9
2.1.1.5. Equations for Confidence and Prediction Intervals.....	12
2.2. Challenge in Estimating Savings Uncertainty	13
2.3. Literature Review.....	13
2.3.1. ASHRAE Guideline 14-2014, <i>Measurement and Verification of Energy, Demand, and Water Savings</i>	13
2.3.2. Uncertainty of “Measured” Energy Savings from Statistical Baseline Models, HVAC&R Research, January 2000. T. Agami Reddy, Ph.D. and David E. Claridge, Ph.D.....	14
2.3.3. Analysis and Improvement on the Estimation of Building Energy Savings Uncertainty, ASHRAE preprint provided by the author, 2012. Yifu Sun and Juan-Carlos Baltazar, Ph.D.	15
2.3.4. International Performance Measurement and Verification Protocol Volume 1.....	15
2.3.5. FEMP M&V Guidelines: Measurement and Verification for Performance-Based Contracts Version 4.0	16
2.3.6. <i>Regression for M&V: Reference Guide</i> , Bonneville Power Administration, 2012.	16
2.3.7. <i>Verification by Energy Modeling Protocol</i> , Bonneville Power Administration, 2012.....	17
2.3.8. <i>Applied Data Analysis and Modeling for Energy Engineers and Scientists</i> , Springer (2011). T. Agami Reddy.....	17
2.3.9. Bayesian Analysis of Savings from Retrofit Projects, found on the web. Paper was to be published in ASHRAE Transactions, Volume 118, Part 2. 2012. John Shonder and Piljae IM.	17
2.3.10. Notes and Memos.....	18
2.3.11. Other Published Papers.....	19
3. DESCRIPTION OF METHODS ANALYZED •(UNCERTAINTY EQUATIONS AND CROSS-CHECKS)	20
3.1. ASHRAE Guideline 14 Equation	20
3.2. Improved Approach Based on Guideline 14	21
3.3. Exact Formula for Ordinary Least Squares Regression.....	21
3.3.1. Derivation	21
3.3.2. Discussion.....	24
3.3.3. Handling Autocorrelation.....	25
3.3.4. Development from Confidence and Prediction Interval Equations.....	25

3.4. Bootstrap Approaches	26
3.4.1. Bootstrap Approach Assuming All Values are Independent.....	29
3.4.1.1. Data Set.....	29
3.4.1.2. Bootstrap Process	30
3.4.2. Block Bootstrap Approach for Autocorrelated Residuals.....	34
3.4.3. Bootstrap Approach for Data With a Relationship Between Independent Variable Values	36
4. RESULTS FOR EACH METHOD ANALYZED	39
4.1. Results for Synthetic Data with a Linear Relationship and No Autocorrelation.....	39
4.1.1. Coefficients	41
4.1.2. Uncertainty.....	41
4.2. Results for Synthetic Data with a Linear Relationship and Moderate Autocorrelation	44
4.3. Results for Synthetic Data with a Linear Relationship, Higher Scatter and Higher Autocorrelation.....	46
4.4. Results for Real Data with a 4-Parameter Relationship, X-Values Not Independent.....	47
5. CHOSEN APPROACH FOR UNCERTAINTY ESTIMATION IN ECAM	49
6. RECOMMENDED FURTHER WORK.....	50

1. EXECUTIVE SUMMARY

1.1. Reason for the Work

This report documents the development and testing of approaches to estimating the uncertainty of savings estimates based on regression models. Some of the approaches are applicable to other types of data-driven models besides regression.

The analysis of the various approaches was intended to confirm the validity of the implementation of one particular approach implemented in the Energy Charting and Metrics tool (ECAM). However, the analyses went further and estimated the validity or consistency of the various approaches tested. This report is intended to not only describe the ECAM approach, but also to make recommendations for further work, and to be educational in nature for an audience not deeply-versed in the statistics of ordinary linear regression.

This work is important in the context of changes in the energy efficiency industry. The reliance on whole building programs is increasing. Such programs include existing building commissioning and building tune-ups, strategic energy management, and pay-for-performance. In all cases, it is desirable to not only have good estimates of energy savings, but to understand the uncertainty in those estimates. It is also of increasing importance to credibly estimate demand savings, whether from these same programs, or from demand response programs.

1.2. Summary of the Work

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
4. Real data, 4-parameter relationship, X-values not independent

The first data set met all of the requirements for ordinary least squares linear regression, and was used to check all of the methods.

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

- 0.0.
 1. Algebraic solution for aggregated uncertainty from OLS regressions, based on a derivation from Josh Rushton, Ph.D. 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, from ASHRAE Guideline 14
 3. Improved FSU, from Yifu Sun and Juan-Carlos Baltazar, Ph.D.

4. Bootstrap Resampling

- a. Resample Data X-Y Pairs
- b. Resample Residuals
- c. Resample Normal Residuals

The results showed that the OLS approach, improved FSU, and Bootstrapping X-Y pairs provide almost identical results for a linear data set without autocorrelation. Their results match regardless of the length of the reporting period. ASHRAE FSU is close, but deviates for reporting periods less than or greater than about 6 to 7 months.

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

For data where the X-values are related, such as energy models based on outside air temperature, the approaches based on linear methods (OLS and FSU) appear to overestimate the uncertainty. The OLS approach is close, but the Improved FSU approach overestimated the uncertainty by more than 30% for data set 4. 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 only 12% 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 formulas for estimating uncertainty based on OLS seemed to work fairly well even for data requiring a 4-parameter model.

The biggest issue appears to be for data sets with autocorrelation. The simple autocorrelation adjustment from ASHRAE is designed for models based on daily data. Hourly data has much more autocorrelation, and the autocorrelation may cover many lags—energy use may have a correlation to not only the energy use one hour prior, but several hours prior, a day prior, and even a week or more prior. Therefore, the algebraic approaches to estimating the impact of autocorrelation should not be trusted if applied to hourly data.

1.3. Author's Comment

Because of these industry changes, there also 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 my opinion, it is valuable for both engineers and statisticians to learn from each other. 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, I believe that they should usually have physical significance.

Similarly, 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, overall, this work may then, to some extent, reflect my opinions regarding the type of information that is valuable, but I believe it could be of wide interest in the industry, and particularly to energy engineers involved.

2. BACKGROUND AND LITERATURE REVIEW

This section describes foundations of uncertainty estimation for ordinary least squares regressions (OLS), the challenges and issues in applying those approaches for energy models, and reviews some of the literature used in the energy efficiency industry and in preparing these analyses.

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.

2.1. Foundations of Models and Uncertainty

Engineers’ estimation of uncertainty, when estimated, have likely thus far been based on the ASHRAE Guideline 14-2014, *Measurement and Verification of Energy, Demand, and Water Savings*, or its predecessor, ASHRAE Guideline 14-2002, *Measurement and Verification of Energy and Demand Savings*. The approach is described in the Guideline’s Annex B, “Determination of Savings Uncertainty.” However, the Annex does not thoroughly explain the basis for the calculations. This section provides further explanation. Section 2.2 provides references for the information in Guideline 14, as well as other documents pertaining to the estimation of savings uncertainty of data-driven models.

2.1.1. Confidence and Prediction Intervals for Ordinary Least Squares Regressions

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 context of regression:

- An Estimate is the average, or expected, y-value, given a specific x-value. The uncertainty in a regression Estimate is a Confidence Interval.

- A Prediction is the specific y-value that may accompany a specific x-value. The uncertainty in a regression Prediction is a Prediction Interval.

The calculations of uncertainty in a model estimate or a point prediction is easily found in statistical literature. The BPA *Regression for M&V: Reference Guide*¹ provides an explanation of Confidence Levels, Confidence Intervals, and Prediction Intervals. To help this document stand alone, part of that information is excerpted or paraphrased here.

2.1.1.1. Confidence Level

“Uncertainty is associated with a given confidence level or probability – for example, ‘We are 90% confident that the range 433 and 511 kWh bands the true value,’ or, as it is more commonly but less accurately expressed, ‘We are 90% confident that the true value lies between 433 and 511 kWh.’ Confidence level is an input number; for a given sample and regression, the higher the confidence level specified, the larger the estimated range that is likely to contain the true value that proportion of the time.”

2.1.1.2. Confidence Interval

“Confidence intervals define the range – an uncertainty band – that is expected to band the true regression, with a certain probability. The width of the confidence interval provides some idea of uncertainty about the estimated parameters. For example, the results of a regression analysis of savings may be reported as ‘500 kWh \pm 5% at the 95% confidence level’, this means that there is a 95% chance that the confidence interval of 475 to 525 kWh contains the true value of savings. A statement of ‘500 kWh \pm 5% at the 68% confidence level’ means that there is only a 68% chance that the true savings value is between these calculated limits, and a 32% chance that it is outside them.”

Figure 1-1 shows a 95% confidence interval for a regression line. Contrast this with Figure 1-2, which shows 80% confidence intervals for the same set of data. The 95% confidence boundary is wider than the 80% confidence boundary. (Because the interval is quite narrow near the center of the data, it may appear that the boundary lines cross the black line for the model. They do not.) The confidence intervals in Figures 1-1 and 1-2 demonstrate that higher probability requirements (e.g. 95%) that an interval contains the true regression line require wider intervals than lower probabilities (e.g. 80%.) The wider the confidence interval, the more likely it is to contain the true regression line for the entire population of points.

¹ The BPA *Regression for M&V: Reference Guide* is available at <https://www.bpa.gov/EE/Policy/IManual/Pages/IM-Document-Library.aspx>

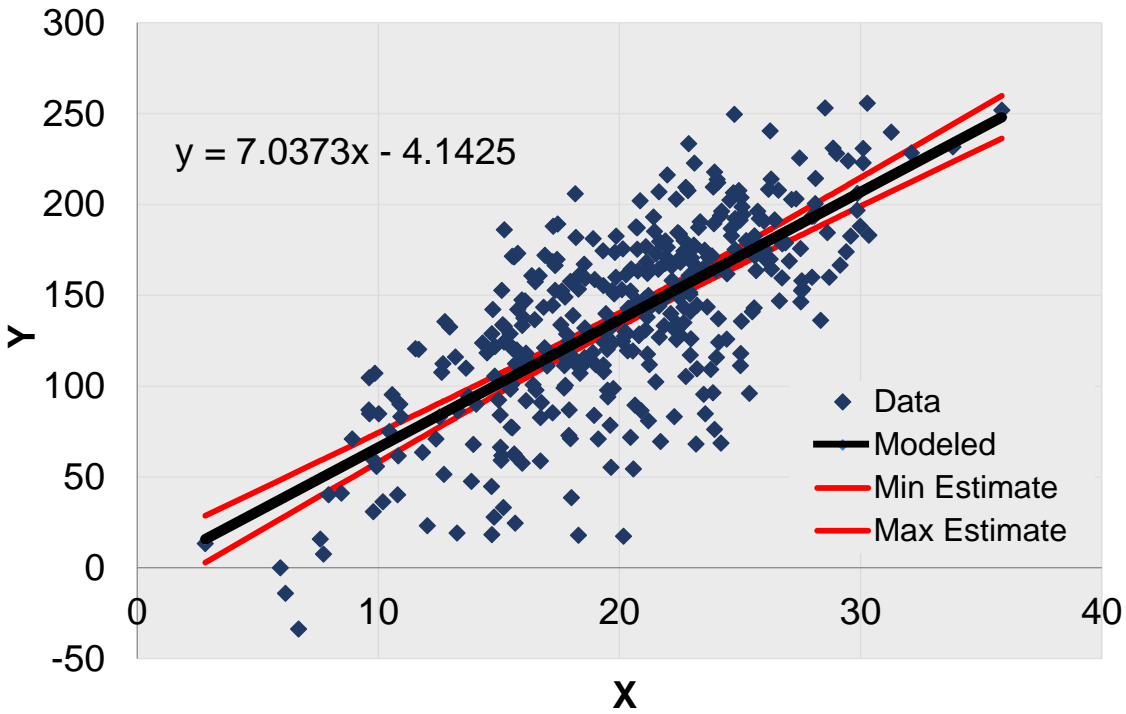


Figure 1-1. Data and Regression Line with a 95% Confidence Interval

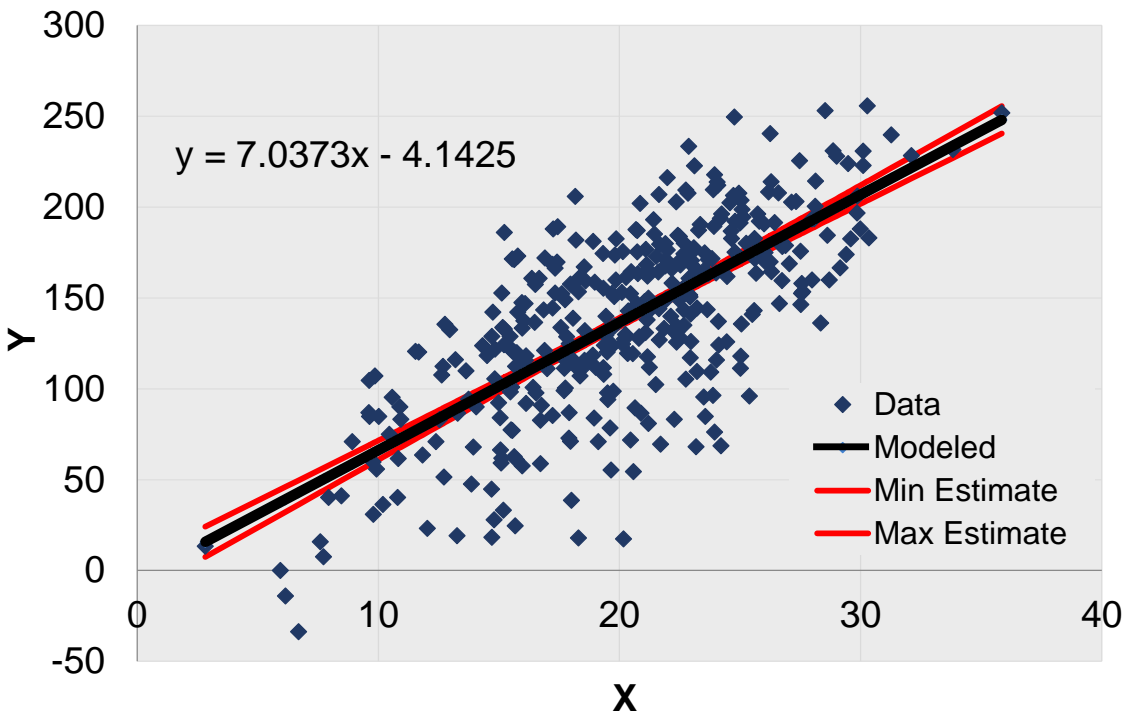


Figure 1-2 . Data and Regression Line with an 80% Confidence Interval

2.1.1.3. Prediction Interval

A prediction interval is an estimate, based on earlier observations, of the interval in which future data points will fall, for an input probability. Prediction intervals are similar to confidence intervals, but rather than estimating the distribution of a parameter, prediction intervals are used to predict the distribution of future points by looking at the distribution of prior points.

The prediction intervals in Figures 1-3 and 1-4, using the same data set as for Figures 1-3 and 1-4, demonstrate that higher probabilities that an interval contains the true regression line require wider intervals than lower probabilities. In other words, the wider the confidence interval, the more likely it is to contain a future point. If you could count the points in Figure 1-3, you would find that 95% of them fall within the prediction interval. Similar, for Figure 1-4, 80% of the points fall within the prediction interval.

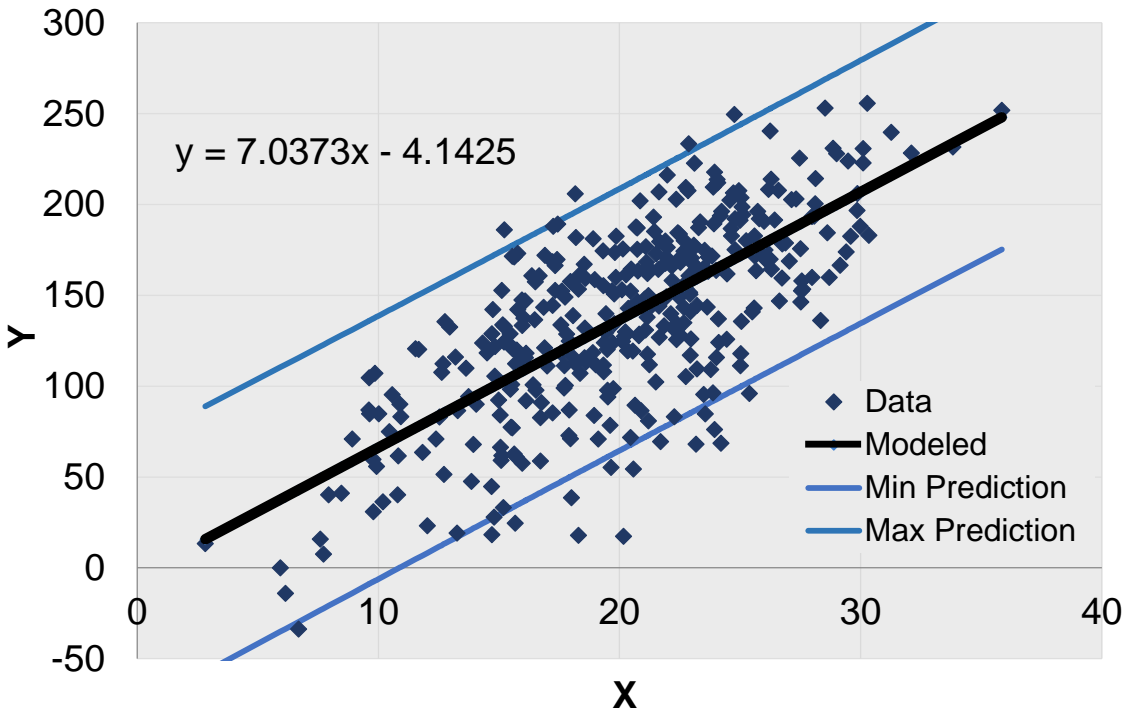


Figure 1-3. Data and Regression Line with a 95% Prediction Interval

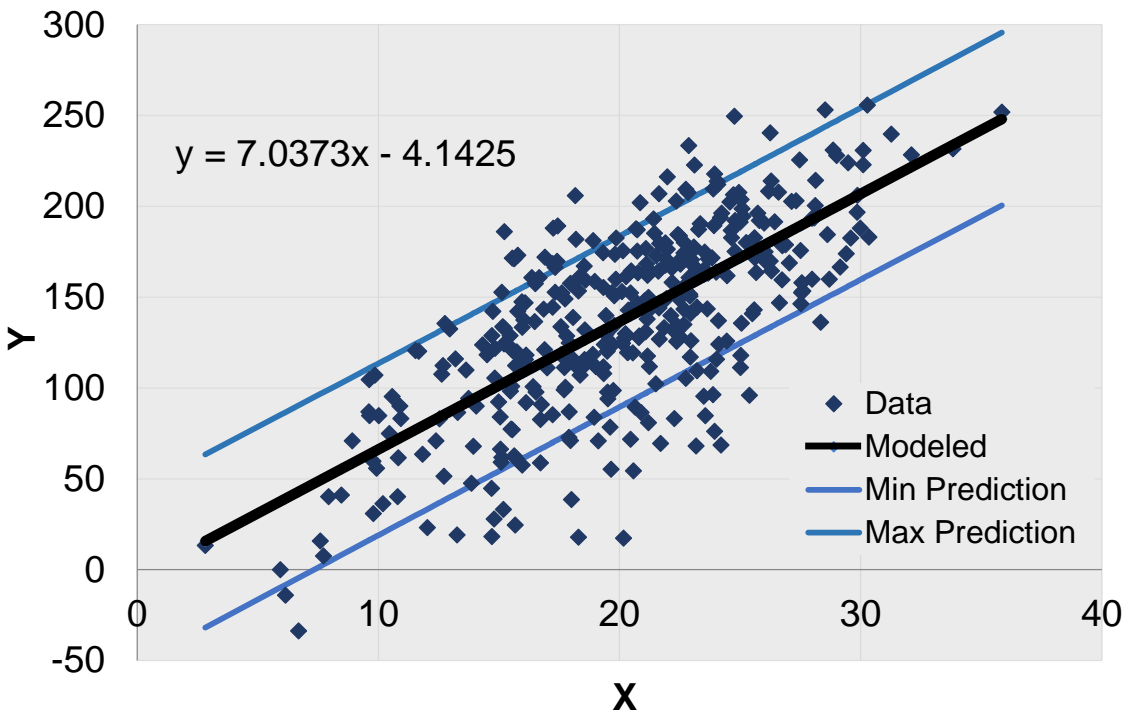


Figure 1-4. Data and Regression Line with an 80% Prediction Interval

2.1.1.4. Putting it All Together

There are two separate types of uncertainty that need to be accounted for in estimating the uncertainty in an adjusted baseline, and these two types of uncertainty are related to the confidence and prediction intervals. Gaining an understanding of confidence and prediction intervals will help lead to an understanding of the uncertainty in the total uncertainty associated with the predictions for many points (x-y pairs) such as for an adjusted baseline.

Figure 1-5 shows both confidence and prediction intervals for the two different confidence levels shown above, 80% and 95%. This is a duplicate of Figure 4-4 of the BPA Regression for M&V: Reference Guide.

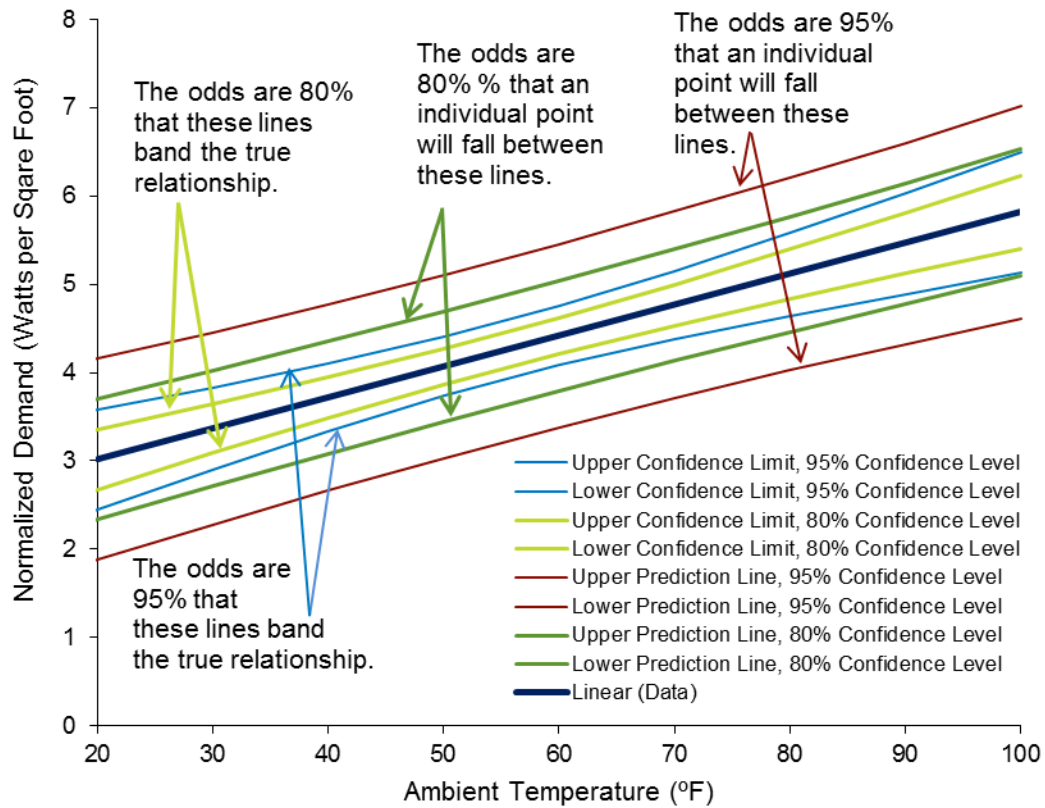


Figure 1-5 Data, Regression Line, and 95% Confidence and Prediction Intervals

Figure 1-6 shows both confidence and prediction intervals for a 95% confidence level.

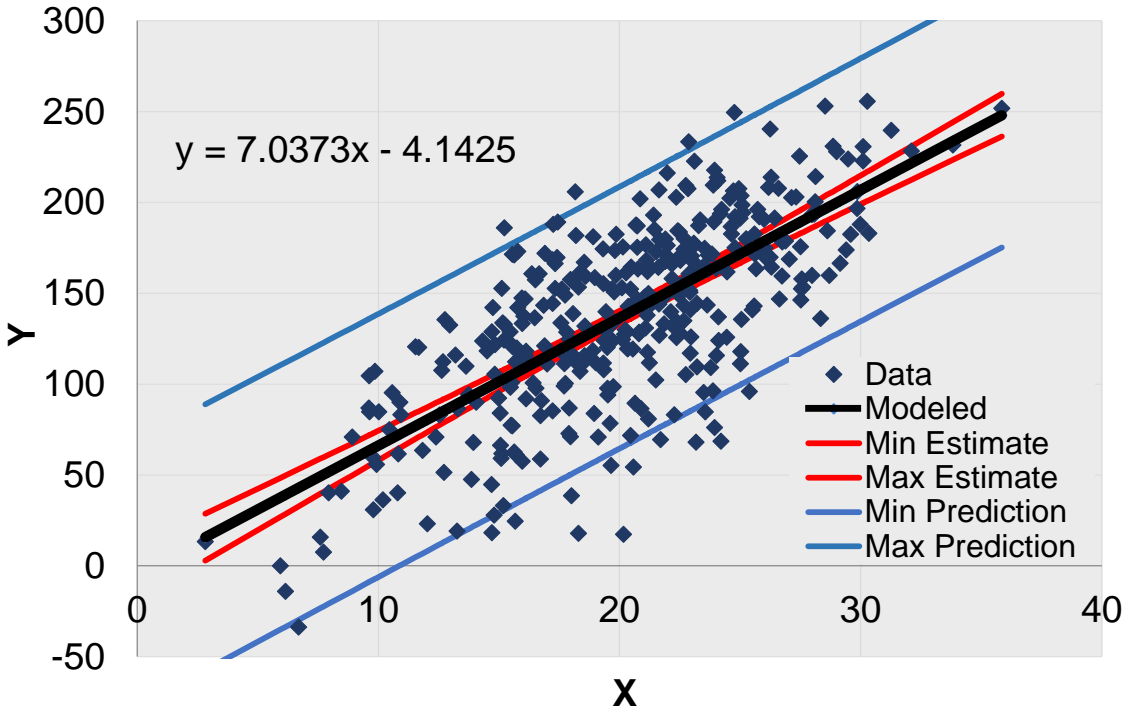


Figure 1-6 Data, Regression Line, and 95% Confidence and Prediction Intervals

Figure 1-7 “zooms” into the lower left part of this to further see what this means in terms of uncertainty:

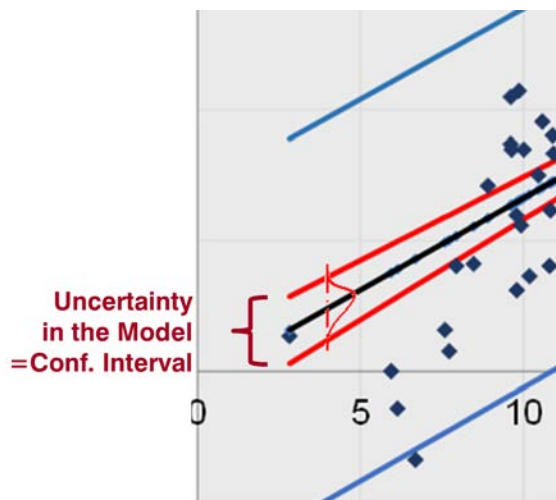


Figure 1-7. Detail of Model Confidence Interval

The confidence interval is the uncertainty in the model, i.e. in the regression line. Furthermore, the uncertainty is assumed to be a normal distribution based on the ordinary linear regression assumption that the deviations of the points from the regression line are a normal distribution.

There is additional uncertainty, beyond the model uncertainty represented by the Confidence Interval, to be considered in making the prediction for an individual point, as shown in Figures 1-8 and 1-9. Figure 1-8 shows the additional variability, and that it is the t-statistic times the standard error, or root mean squared error. From the BPA *Regression for M&V: Reference Guide*:

“Root mean squared error (RMSE) is an indicator of the scatter, or random variability, in the data, and hence is an average of how much an actual y-value differs from the predicted y-value. It is the standard deviation of errors of prediction about the regression line. Standard error of the estimate (SE) is always adjusted by the number of parameters in the model. Keep in mind, however, that some sources include the adjustment for the number of parameters in their definition of RMSE; others do not. In this document, SE and RMSE are synonymous, and include the adjustment for the number of parameters in the model. Standard error of the estimate is sometimes called standard error of prediction.”

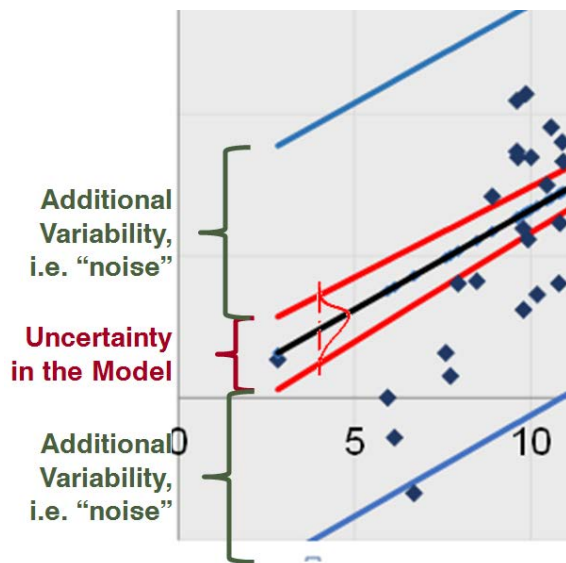


Figure 1-8. Sources of Uncertainty for Regression Predictions

Figure 1-9 illustrates that aggregating the standard errors (multiplied by the t-statistic) with the confidence interval gives us the prediction interval.

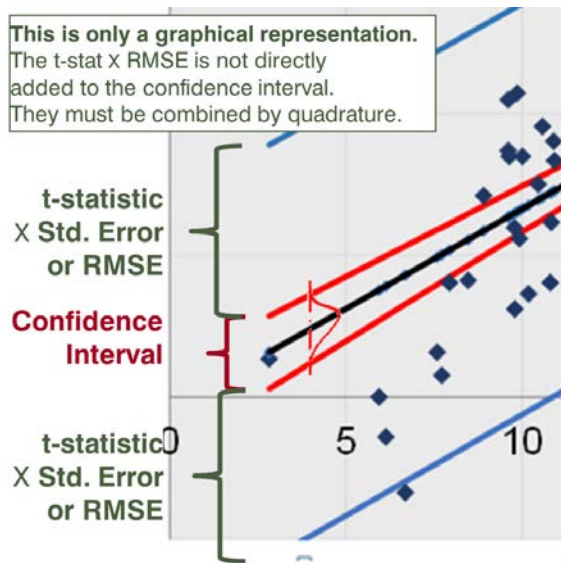


Figure 1-9. Prediction Interval for a Regression

This aggregation is done by quadrature, the square root of the sum of the squares:

$$\text{Prediction Interval} = \sqrt{(t \times \text{Std Error} \times 2)^2 + (\text{Confidence Interval})^2}$$

2.1.1.5. Equations for Confidence and Prediction Intervals

One Side of the Confidence Interval

$$= t \times \text{Std Error} \times \sqrt{\frac{1}{n} + \frac{(x - \bar{x})^2}{\sum_{i=0}^n (x_i - \bar{x})^2}}$$

One Side of the Prediction Interval

$$= t \times \text{Std Error} \times \sqrt{1 + \frac{1}{n} + \frac{(x - \bar{x})^2}{\sum_{i=0}^n (x_i - \bar{x})^2}}$$

Where

- t is the t-statistic for the input confidence level and the number of points in the regression data
- StdError , for Ordinary Linear Regression, is the same as the RMSE of the model residuals.
- n is the number of points in the regression data
- x is the independent variable value for which the interval is being computed
- X_{avg} is the average of all the independent variable values for the points in the regression data

These equations can be found in any reputable source on ordinary linear regression. Note that the only difference between the two equations is the presence of a “1 +” term inside the parentheses. Algebraically, this “1 +” term represents the standard error.

Since the uncertainty in the model—the confidence interval—and the standard error (times the t-statistic) are independent of each other, they are combined by quadrature to get the total prediction interval. This may seem counterintuitive, since both of these terms include the standard error. However, the uncertainty in the model means that the model could be slightly higher or lower than the true relationship, and/or have a different slope. But if the model is different from the true relationship, it is different in the same way for every single prediction it makes. In contrast, the standard error represents “noise,” and it can be positive or negative, and have a different value, for every prediction. This will be a key to further understanding.

2.2. Challenge in Estimating Savings Uncertainty

So, if all of this is well known, and described in statistical literature, why has the industry seemingly been challenged to properly use uncertainty in estimating savings from regressions?

There are at least two important reasons:

- 0.0. 1. Uncertainty in savings estimates requires the aggregation of multiple predictions. There is a baseline prediction and associated uncertainty for every point in a reporting period. What is the combined uncertainty associated with all of these baseline predictions? The answer to this seems to be much harder to find in the literature.
2. Many of the models used for estimating baseline energy use, and hence savings, violate the requirements for ordinary linear regression. See the BPA Regression for M&V: Reference Guide for more information on these requirements.

This document covers both of these issues; the first in a comprehensive fashion, and the second with some examples using resampling methods to check the validity of the approach based on ordinary linear regression.

2.3. Literature Review

This section describes various documents that were reviewed during the development and analysis of the various methods. It includes documents that are foundational or provide important background.

2.3.1. ASHRAE Guideline 14-2014, *Measurement and Verification of Energy, Demand, and Water Savings*

This covers uncertainty of regression-based energy savings in [Annex B: Determination of Savings Uncertainty](#), especially in Section B4 “Uncertainty of Regression-Based Savings Models.”

It says that for utility bills, the measurement errors can be neglected, and the total uncertainty is based on the model prediction errors. It mentions three specific cases to be considered:

- 0.0. 1. weather-independent models, when for example lighting retrofits are being evaluated,
2. weather-based regression models with uncorrelated residuals as assumed when analyzing monthly utility bills, and
3. weather-based regression models with serially correlated residuals, as is often encountered with models based on hourly or daily data.

Focusing on the second case for now, Annex B describes the fractional savings uncertainty—the uncertainty divided by the savings. This is closely related to statistical precision. Then it says,

“Reddy and Claridge (2000) presented a method for determining uncertainty in actual savings for cases where baseline energy use can be fit to a linear model dependent on weather and/or other variables.” The Annex proceeds to show the matrix algebra to estimate uncertainty, and then says, “Reddy and Claridge then developed an approximation to eliminate the need for matrix algebra.”

The approximation uses a 1.26 factor, which was empirically derived as described in Reddy and Claridge 2000. It also includes the assumption that there is no uncertainty in any estimated change points.

Focusing on the third case, weather-based regression models with serially correlated residuals, Annex B documents another recommendation by Reddy and Claridge. They recommended the use of the autocorrelation coefficient ρ , assuming only lag-1 autocorrelation, to adjust the number of data points down to an effective number of *independent* data points. This number of independent data points is substituted for the number of actual data points in several places in the approximate equation.

They also provide a short discussion of the calculation of ρ and its applicability for different data intervals.

2.3.2. Uncertainty of “Measured” Energy Savings from Statistical Baseline Models, HVAC&R Research, January 2000. T. Agami Reddy, Ph.D. and David E. Claridge, Ph.D.

This paper is a foundation for the uncertainty approaches and calculations in ASHRAE Guideline 14. It discusses the statistical criteria for goodness-of-fit, including the coefficient of determination, R^2 , and the coefficient of variation of the root mean square error, $CV(RMSE)$. They discuss how both of those indices are normalized, with R^2 describing how much of the variation is explained by the model, and $CV(RMSE)$ describing how much of the variation is not explained by the model. More significantly, they discuss that R^2 is of lesser utility when slopes are very high or when near zero, but that $CV(RMSE)$ is not sensitive to slope.

Perhaps the most important part of the paper is the derivation of the simplified equation described in Guideline 14 Annex B. They start with a key reference: “The prediction error of a

sum of m future observations (as is needed for determining energy savings) is given by Theil (1971).² The equation from Theil uses matrix algebra and Reddy and Claridge simplified it using numerical trials, arriving at the equation with the 1.26 factor which they said “holds to within about 10% accuracy.”

They also recommend the use of the autocorrelation coefficient ρ to adjust the number of data points down to an effective number of *independent* data points as mentioned in the discussion for ASHRAE Guideline 14.

2.3.3. Analysis and Improvement on the Estimation of Building Energy Savings Uncertainty, ASHRAE preprint provided by the author, 2012. Yifu Sun and Juan-Carlos Baltazar, Ph.D.

This paper describes an improvement to the ASHRAE equation for fractional savings uncertainty developed by T. Agami Reddy, Ph.D. and others, and published in *Uncertainty of “Measured” Energy Savings from Statistical Baseline Models* and in ASHRAE Guideline 14.

From the paper’s introduction: “In this paper, the derivation of the original matrix algebra expression for energy savings uncertainty and Reddy and Claridge’s simplified expression is studied for the case of weather dependent model. Then, the comprehensive equation and the simplified equation are compared to analyze their relationships, which are verified by a series of cases. Based on the case study, how these assumptions impact on the uncertainty estimation is illustrated. Eventually, an improved method is introduced to calculate the energy savings uncertainty both simply and accurately, excluded the impact from the different number of monitored month under the savings analysis period.”

Baltazar and Sun show that the ASHRAE simplified method, with the 1.26 factor, is correct for a certain number of post-implementation time periods, but deviates for longer and shorter reporting periods. They introduced an improved method that replaces the 1.26 factor with a second order polynomial.

2.3.4. International Performance Measurement and Verification Protocol Volume 1

IPMVP 2012, *Concepts and Options for Determining Energy and Water Savings Volume 1* is the prior international standard for M&V. It has now been superseded by International Performance Measurement and Verification Protocol *Core Concepts*, April 2016. The Core Concepts are intended to be supplemented by “Application Guides.” However, the Application Guide for Statistics and Uncertainty has yet to be finalized, so this discussion refers to *Statistics and Uncertainty for IPMVP*, EVO 2014. Appendix B Uncertainty of the 2012 Volume I document is very similar.

² Thiel, H. 1971. *Principles of Econometrics*. John Wiley & Sons, New York.

Statistics and Uncertainty for IPMVP covers sources of uncertainty, defines statistical terms, and covers issues and calculations regarding modeling, sampling, and metering. It then discusses combining components of uncertainty, provides calculations of uncertainty for point estimates and for savings, and gives an example.

In my opinion, the formula provided for aggregating multiple point prediction uncertainties into an aggregate uncertainty is flawed. It does not provide an empirical factor to simplify matrix algebra like ASHRAE, but assumes that aggregate uncertainty, e.g. the standard error of the annual savings estimate, is simply the square root of the sum of the squares of the standard error of the estimate. This will underestimate the uncertainty, because it does not include the uncertainty in the model itself.

2.3.5. FEMP M&V Guidelines: Measurement and Verification for Performance-Based Contracts Version 4.0

The FEMP M&V Guidelines do not cover the uncertainty associated with regression models, but refer to Guideline 14 and IPMVP.

2.3.6. Regression for M&V: Reference Guide, Bonneville Power Administration, 2012.

Part of the Implementation Manual Document Library available as of March 2017 at <https://www.bpa.gov/EE/Policy/IManual/Pages/IM-Document-Library.aspx>

This document is intended as a primer on ordinary linear regression for energy engineers and analysts. It provides information on the advantages and disadvantages of regression and describes a recommended regression process. It introduces the common model types used, primarily for commercial and residential buildings, and discusses the use of continuous and categorical variables.

Confidence levels, confidence intervals, prediction intervals, and their relationships to savings uncertainty are explained in layperson's or engineer's terms, while trying to be faithful to the accurate statistical meaning.

The process of validating models, including an explanation of statistical tests and measures, is covered along with the formulas to calculate the statistical tests in Microsoft Excel[®]. Model bias, including how its definitions and descriptions vary in the energy efficiency literature, is described. Analysis of residuals is introduced, but it does not include the very useful time series plot of residuals. The document concludes with an example and an Appendix with a glossary of statistical terms.

The document does not cover the aggregate uncertainty of many predictions. It gives an example, but it is not clear how the aggregate uncertainty is derived.

Full disclosure: I was the original primary author of this document.

2.3.7. Verification by Energy Modeling Protocol, Bonneville Power Administration, 2012.

Part of the Implementation Manual Document Library available as of March 2017 at <https://www.bpa.gov/EE/Policy/IManual/Pages/IM-Document-Library.aspx>

The Energy Modeling Protocol overlaps some with the Regression Reference Guide, but it is intended as an application document, more prescriptive and less educational. It describes using regression or other data-driven models in general, then goes into procedures. It describes the use of charts, identifying measurement boundaries, selecting time intervals, and calculating savings. A unique but very useful section contrasts multiple regression and multiple models, tying the multiple model approach into the distinction between continuous and categorical variables made in the Regression Reference Guide.

The document recommends the use of the ASHRAE Fractional Savings Uncertainty equations. It discusses tools available for modeling and estimating savings, and provides an example.

Full disclosure: I was the original primary author of this document.

2.3.8. Applied Data Analysis and Modeling for Energy Engineers and Scientists, Springer (2011). T. Agami Reddy

This textbook covers the wide range of topics implied by its title. Topic areas include probability concepts, data collection and preliminary data analysis, statistical inferences, least squares regression, design of experiments, optimization methods, data mining methods, analysis of time series data, parameter estimation methods, inverse methods, and risk analysis.

For the work discussed in this paper, sections dealing with ordinary linear regression and time series data analysis were used for learning and review. In particular, the sections covering the Bootstrap and other Resampling Methods were studied. Several other sections, such as “Analysis of Time Series Data,” were skimmed for relevant information and understanding.

2.3.9. Bayesian Analysis of Savings from Retrofit Projects, found on the web. Paper was to be published in ASHRAE Transactions, Volume 118, Part 2. 2012. John Shonder and Piljae IM.

From the abstract, “The objective of this paper is to show that Bayesian inference provides a consistent framework for estimating savings and savings uncertainty in retrofit projects. The paper discusses prior methods to estimate savings uncertainty from regression models, including the method in ASHRAE Guideline 14. It states that the prior methods are approximations, and have limitations and inaccuracies.

This paper covers two major examples, the first data set was fit using a simple linear model having no autocorrelation, and the second was fit using a 5-parameter model. The second data set also had significant autocorrelation. For the first example, the paper shows that the

Bayesian approach used provides the same result as frequentist statistics, for both the mean response estimate and for the uncertainty. For the second example, the Bayesian approach provided an estimate of savings that was 5% lower than the 5-parameter model, and a 95% confidence interval that was 37% wider. However, I viewed these results with suspicion. The paper only showed the 5-parameter model plotted with the data, not the Bayesian model. The parameters derived using the Bayesian approach, when plotted, appeared to be a much poorer fit to the data.

Although there is a possible issue with the second example, the paper was useful in describing a resampling approach that is related to the Bootstrap approach described later in this paper.

2.3.10. Notes and Memos

Jim Stewart, Ph.D. of Cadmus wrote a memo, dated August 24, 2016, as part of an industrial program impact evaluation. The subject was a *Proposed Framework for Estimating Standard Error of Forecast Method Energy Savings*. The memo described a “recommended approach for estimating the standard errors of forecast model savings estimates. It presented “a framework for deriving the standard error of the forecast model savings estimates,” provided an example, and presented formulas for standard errors of both forecast and pre-post type models.

In summary, the memo “demonstrated that forecast savings estimate has two sources of uncertainty: the first is the variance of the adjusted baseline and the second is the variance of metered energy use. Both components should be accounted for to obtain an accurate estimate of the variance of the savings estimate.” This is the same approach that was derived by Josh Rushton and described in Section 3. However, this memo retains the matrix algebra for of the calculation.

The following class notes provide some confirmation of the ASHRAE approach for correlated residuals.

http://seismo.berkeley.edu/~kirchner/eps_120/Toolkits/Toolkit_11.pdf

This is part of the “condensed notes” for a class on Analysis of Environmental Data at the University of California, Berkeley:

http://seismo.berkeley.edu/~kirchner/eps_120/EPSToolkits.htm

The notes describe how to “Adjust your uncertainty estimates to account for the loss of degrees of freedom. Serial correlation leads to underestimates of uncertainty (and thus overestimates of statistical significance) because when residuals are not independent, the effective number of degrees of freedom can be much smaller than the sample size would indicate. In principle, one should be able to account for this loss of degrees of freedom by calculating the “effective” sample size—that is, the number of independent measurements would be functionally equivalent to your set of nonindependent measurements. You then use this “effective” sample size, n_{eff} , in place of n in calculating standard errors and performing significance tests.”

The notes provide the same equation to adjust the number of data points using the lag-1 autocorrelation coefficient. It also describes issues with this approach, including that it assumes

that ρ is the true autocorrelation coefficient, even though it is based on a sample, and provides a reference for a proposed improvement.

Another confirmation of the ASHRAE approach for correlated residuals was found in some notes for a class at the University of Arizona, "GEOS 585A, Spring 2015." The notes themselves appear to originally be from Georgia Tech, but I could not find the originals. The title of the notes is simply "Autocorrelation." The last section in the notes is titled "Effective Sample Size" and it provides the same equation as Reddy and Claridge to scale the number of data points down to the number of "effective" data points.

It seems certain that a text on econometrics, time series analysis, or regression would present the same equation, but I did not review any complete textbooks on these subjects.

2.3.11. Other Published Papers

The papers listed below are either predecessors to documents listed above, or useful background information on the uncertainty of energy savings estimates from data-driven models.

Reddy, T. A., J. K. Kissock, and D.K. Ruch, 1998. "Uncertainty in Baseline Regression Modeling and in Determination of Retrofit Savings" *Journal of Solar Energy Engineering*.

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

Subbarao, Kris, Lei Yafeng, and T. Agami Reddy, 2011. "The Nearest Neighborhood Method to Improve Uncertainty Estimates in Statistical Building Energy Models," ASHRAE Preprint. To be published in *ASHRAE Transactions*, Volume 117, Part 2.

3. DESCRIPTION OF METHODS ANALYZED (UNCERTAINTY EQUATIONS AND CROSS-CHECKS)

This section introduces the formulas and/or approaches for each of the methods analyzed.

3.1. ASHRAE Guideline 14 Equation

The following is paraphrased from the BPA Energy Modeling Protocol, although it originally came from Guideline 14. The form of the simplified ASHRAE equation for “weather models with correlated residuals “can cover cases with and without autocorrelation, so that is the form shown here:

Weather models with correlated residuals are models where each point has a relationship with the points associated with recent prior timestamps. There is the potential for correlated residuals (known as time-series autocorrelation) when the time unit is short, such as with hourly models. There can also be autocorrelation with daily models.

For these types of models, the equation is:

$$\frac{\Delta E_{save,m}}{E_{save,m}} = t \times \frac{1.26 \cdot CV \left[\frac{n}{n'} \left(1 + \frac{2}{n'} \right) \frac{1}{m} \right]^{1/2}}{F}$$

Where:

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

$$CV(RMSE) = \frac{\sqrt{[\sum_i \times (E_i - \hat{E}_i)^2 / (n - p)]}}{\bar{E}}$$

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

t=Student’s t-statistic

()_i = measured value

(^)_i = predicted value

(_)_i = average value

n= number of points in the baseline period

m= number of points in the post period

p=the number of model parameters

n' is the effective number of points after accounting for autocorrelation.

$$n' = n \cdot \frac{1 - \rho}{1 + \rho}$$

Where:

ρ is the autocorrelation coefficient, which is the square root of the R² calculated for the correlation between the residuals and the residuals for the prior time period.

3.2. Improved Approach Based on Guideline 14

This is the improved formula developed by Sun and Baltazar.

They showed that the deviation of the simplified equation results from the true value based on the comprehensive matrix equation was in the form of a polynomial. Therefore, the improved method 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 has daily or monthly intervals.

The improved formula is:

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

Where:

M=number of months of reporting period data
a, *b*, and *c* are defined as follows:

Data Interval	Monthly	Daily
<i>a</i>	-0.00022	-0.00024
<i>b</i>	0.03306	0.03535
<i>c</i>	0.94054	1.00286

3.3. Exact Formula for Ordinary Least Squares Regression

3.3.1. Derivation

Josh Rushton, Ph.D. (Mathematics and Stochastic Processes, University of Wisconsin-Madison), provided a derivation of the aggregated uncertainty of many predictions. This derivation is provided below, with permission. It is derived for OLS regressions. Part of the purpose of work for this paper was to test the derived formula's applicability to more complex regressions.

Assume the model that generates the data is of this form:

$$y = \beta_0 + \beta_1 \cdot x + \varepsilon, \quad \text{where } \varepsilon \sim N(0, \sigma^2)$$

We fit the model using pre-period data $(x_i^{\text{pre}}, y_i^{\text{pre}})_{i=1}^{n^{\text{pre}}}$, and the fitted model is this:

$$y = \hat{\beta}_0 + \hat{\beta}_1 \cdot x + \varepsilon, \quad \text{where } \varepsilon \sim N(0, s^2)$$

We want to use the fitted model to predict total consumption for a post-period data set, $(x_i^{\text{post}})_{i=1}^{m^{\text{post}}}$.

The estimate is this:

$$\hat{y} = \sum_{i=1}^{m^{\text{post}}} (\hat{\beta}_0 + \hat{\beta}_1 x_i^{\text{post}}) = \hat{\beta}_0 \times m^{\text{post}} + \hat{\beta}_1 \times \sum_{i=1}^{m^{\text{post}}} x_i^{\text{post}} \quad (1)$$

This expression does not include error terms, so the only uncertainty embedded in \hat{y} is uncertainty in the estimated parameters $\hat{\beta}_0$ and $\hat{\beta}_1$. We can think of \hat{y} as our fitted model's best guess at the mean total consumption for n observation periods with the temperature values $(x_i^{\text{post}})_{i=1}^{m^{\text{post}}}$.

Model uncertainty (uncertainty in the mean total consumption for the post period):

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

Note that if $m_{\text{post}} = n_{\text{base}}$ and $x_i^{\text{post}} = x_i^{\text{base}}$ (i.e., post period conditions are same as pre-period), the model uncertainty reduces to this:

$$se(\hat{y}) = s \cdot \sqrt{n}$$

Intuitively, the reason is that the data set used to fit the model, $(x_i, y_i)_{i=1}^n$ basically constitutes a single observation of the actual total consumption for a year like that, and the actual total consumption (which is the same as \hat{y} in this case) equals the mean consumption, $E[\hat{y}]$, plus n noise terms.

Derivation

Several steps are needed to derive expression (2) from the standard OLS matrix formulas. First, we first re-write equation (1) in matrix terms:

$$\hat{y} = [m_{\text{post}} \quad \sum x_i^{\text{post}}] \begin{bmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \end{bmatrix} = \mathbf{x}^{\text{post sum}'} \begin{bmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \end{bmatrix} \quad (3)$$

To calculate the variance of \hat{y} we use the fact that for any random vector \mathbf{z} and any compatible non-random matrix M , the covariance of the matrix product is given by $\text{Cov}(M\mathbf{z}) = M\text{Cov}(\mathbf{z})M'$. In the present case, $\mathbf{z} = \hat{\boldsymbol{\beta}} = (\hat{\beta}_0, \hat{\beta}_1)'$ and $M = \mathbf{x}^{\text{post sum}'}$ yields this expression:

$$\begin{aligned} \text{Var}(\hat{y}) &= \mathbf{x}^{\text{post sum}'} \text{Cov}(\hat{\boldsymbol{\beta}}) \mathbf{x}^{\text{post sum}} \\ &= \mathbf{x}^{\text{post sum}'} \sigma^2 (X'X)^{-1} \mathbf{x}^{\text{post sum}} \end{aligned} \quad (4)$$

Here, X is the standard OLS design matrix populated with the baseline (pre-case) data. To put this in a form similar to expression (2), we start with the inner-most part of expression (4) and work our way out (since $(X'X)^{-1}$ only involves pre-case data, we drop the “pre” superscript for the derivation of the inverse):

$$\begin{aligned}
 X'X &= \begin{bmatrix} 1 & 1 & \cdots & 1 \\ x_1 & x_2 & \cdots & x_n \end{bmatrix} \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{bmatrix} \\
 &= \begin{bmatrix} n & \sum x_i \\ \sum x_i & \sum x_i^2 \end{bmatrix} \\
 &= n \begin{bmatrix} 1 & \bar{x} \\ \bar{x} & \bar{x}^2 \end{bmatrix}
 \end{aligned}$$

The inverse of the matrix is product is therefore:

$$\begin{aligned}
 (X'X)^{-1} &= \left(\frac{1}{n}\right) \left(\frac{1}{\bar{x}^2 - (\bar{x})^2}\right) \begin{bmatrix} \bar{x}^2 & -\bar{x} \\ -\bar{x} & 1 \end{bmatrix} \\
 &= \left(\frac{1}{n}\right) \left(\frac{1}{\text{Var}(x)}\right) \begin{bmatrix} \bar{x}^2 & -\bar{x} \\ -\bar{x} & 1 \end{bmatrix}
 \end{aligned}$$

Next, we use the expression for $(X'X)^{-1}$ to re-write (4):

$$\begin{aligned}
 \text{Var}(\hat{y}) &= \mathbf{x}^{\text{post sum}'} \text{Cov}(\hat{\beta}) \mathbf{x}^{\text{post sum}} \\
 &= \mathbf{x}^{\text{post sum}'} \sigma^2 \left(\frac{1}{n_{\text{base}}}\right) \left(\frac{1}{\text{Var}(x_{\text{base}})}\right) \begin{bmatrix} \bar{x}_{\text{base}}^2 & -\bar{x}_{\text{base}} \\ -\bar{x}_{\text{base}} & 1 \end{bmatrix} \mathbf{x}^{\text{post sum}} \\
 &= \left(\frac{\sigma^2}{n_{\text{base}}}\right) \left(\frac{1}{\text{Var}(x_{\text{base}})}\right) [m_{\text{post}}, \sum x_i^{\text{post}}] \begin{bmatrix} \bar{x}_{\text{base}}^2 & -\bar{x}_{\text{base}} \\ -\bar{x}_{\text{base}} & 1 \end{bmatrix} \begin{bmatrix} m_{\text{post}} \\ \sum x_i^{\text{post}} \end{bmatrix} \quad (5) \\
 &= \left(\frac{\sigma^2 m_{\text{post}}^2}{n_{\text{base}}}\right) \left(\frac{1}{\text{Var}(x_{\text{base}})}\right) [1, \bar{x}_{\text{post}}] \begin{bmatrix} \bar{x}_{\text{base}}^2 & -\bar{x}_{\text{base}} \\ -\bar{x}_{\text{base}} & 1 \end{bmatrix} \begin{bmatrix} 1 \\ \bar{x}_{\text{post}} \end{bmatrix}
 \end{aligned}$$

Finally, we multiply out the matrix product in this expression and simplify as follows:

$$\begin{aligned}
 [1, \bar{x}_{\text{post}}] \begin{bmatrix} \bar{x}_{\text{base}}^2 & -\bar{x}_{\text{base}} \\ -\bar{x}_{\text{base}} & 1 \end{bmatrix} \begin{bmatrix} 1 \\ \bar{x}_{\text{post}} \end{bmatrix} &= [\bar{x}_{\text{base}}^2 - \bar{x}_{\text{post}}\bar{x}_{\text{base}}, -\bar{x}_{\text{base}} + \bar{x}_{\text{post}}] \begin{bmatrix} 1 \\ \bar{x}_{\text{post}} \end{bmatrix} \\
 &= \bar{x}_{\text{base}}^2 - \bar{x}_{\text{post}}\bar{x}_{\text{base}} - \bar{x}_{\text{base}}\bar{x}_{\text{post}} + (\bar{x}_{\text{post}})^2 \\
 &= \bar{x}_{\text{base}}^2 - 2\bar{x}_{\text{post}}\bar{x}_{\text{base}} + (\bar{x}_{\text{post}})^2 \quad (6) \\
 &= \text{Var}(x_{\text{base}}) + (\bar{x}_{\text{base}})^2 - 2\bar{x}_{\text{post}}\bar{x}_{\text{base}} + (\bar{x}_{\text{post}}) \\
 &= \text{Var}(x_{\text{base}}) + (\bar{x}_{\text{base}} - \bar{x}_{\text{post}})^2
 \end{aligned}$$

Combining expressions (5) and (6) we obtain

$$\begin{aligned}
 \text{Var}(\hat{y}) &= \mathbf{x}^{\text{post sum}'} \text{Cov}(\hat{\boldsymbol{\beta}}) \mathbf{x}^{\text{post sum}} \\
 &= \left(\frac{\sigma^2 m_{\text{post}}^2}{n_{\text{base}}} \right) \left(\frac{1}{\text{Var}(x_{\text{base}})} \right) \left(\text{Var}(x_{\text{base}}) + (\bar{x}_{\text{base}} - \bar{x}_{\text{post}})^2 \right) \\
 &= \left(\frac{\sigma^2 m_{\text{post}}^2}{n_{\text{base}}} \right) \left(1 + \frac{(\bar{x}_{\text{base}} - \bar{x}_{\text{post}})^2}{\text{Var}(x_{\text{base}})} \right)
 \end{aligned} \tag{7}$$

To translate the variance expression (7) into the standard error, replace σ by s and take the square root. The result is expression (2), shown again below.

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

Noise is simply the baseline standard error aggregated by quadrature for the number of points in the post period:

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

Prediction uncertainty is model uncertainty combined by quadrature with the noise:

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

For this, if $m_{\text{post}} = n_{\text{base}}$ and $x_i^{\text{post}} = x_i^{\text{base}}$, we get:

$$se\left(\hat{y} + \sum \varepsilon_i\right) = \sqrt{n \times s^2 + n \times s^2} = \sqrt{2n} \times s$$

3.3.2. Discussion

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} .

3.3.3. Handling Autocorrelation

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 ASHRAE equation adjusts m_{post} by the ratio of n_{base}/n'_{base} . Empirical checks for the data sets in this work suggest that this adjustment should only be applied to the noise portion, so the final equation is:

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

M&V practitioners needing a simplified method to estimate uncertainty should use this equation.

3.3.4. Development from Confidence and Prediction Interval Equations

This is not a derivation, but just a development of the same equations as above, starting from the confidence and prediction intervals, and using the same aggregations—straight sum or square root of the sum of the squares—as derived above. This shows that the derivation above is consistent with confidence and prediction intervals, and was used to help verify the proper implementation of the above equations in ECAM.

Table 3-1. Development of Model Uncertainty Aggregated Over Many Points

Steps	Standard Confidence half-interval, aggregated by SUM
Confidence half-interval, without t-statistic	$s \times \sqrt{(1/n_{base} + (x - \bar{x})^2 / (\sum (x - \bar{x})^2))}$ (s is the standard error of the estimate for the baseline regression, and is the same as RMSE)
The included adjustment for x different from Baseline period \bar{x}	$+ (x - \bar{x})^2 / (\sum (x - \bar{x})^2)$
Simplified by eliminating adjustment for \bar{x}	$s \times \sqrt{(1/n_{base})}$
Aggregating for m_{post} points	$m_{post} \times s \times \sqrt{(1/n_{base})}$
Algebraic rephrasing	$\frac{s \times m_{post}}{\sqrt{n_{base}}}$

Table 3-2. Development of Standard Error Aggregated Over Many Points

Steps	Standard Error (Noise), aggregated by Quadrature
Standard Error	s
Aggregating by Quadrature (square root of the sum of squares) for m_{post} points	$\sqrt{s^2 \times m_{post}}$
Algebraic rephrasing	$s \times \sqrt{m_{post}}$

3.4. Bootstrap Approaches

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 bootstrap resamples data from a sample to develop parameter distributions and or distributions of results. The key assumption for the bootstrap is that the distribution of the parameter in the sample closely approximates its distribution in the population, and that the distribution in the bootstrap sample (resample) closely approximates its distribution in the sample. I.e. the

bootstrap sample is a close representation of the sample, and the sample is a close approximation of the population.

The bootstrap's uncertainty estimates were accepted 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, was taken as support for a simpler method, and disagreement was taken as cause for concern. Further study is recommended to confirm the reliability of such comparisons.

In a bootstrap, a sample is drawn, with replacement, from the original sample; parameters are estimated; the savings is estimated; and the process is repeated many times. "With replacement" means that after an item is drawn for the bootstrap sample, it is placed back into the original sample 'deck,' so that it is again available to be drawn as part of the bootstrap sample. Therefore, for each bootstrap sample, some items can be drawn multiple times, whereas others may not be drawn at all. Each bootstrap sample, however, has the same number of members as the original sample.

There are various criteria to determine how many bootstrap samples are needed, but the approach I used was to simply plot the distributions of the savings estimates until the distribution was reasonably smooth and would not change significantly with additional bootstrap samples.

Because of my familiarity with Microsoft Excel®, and to provide transparency of the work, all of these analyses were done in Excel®.

It was also very important to the flexibility needed for these comparisons: For most of these tests two separate draws were completed for each savings estimate. This was done to get correspondence with the two sources of uncertainty described in Section 2:

- 0.0.
 1. Model Uncertainty
 2. Noise

Getting these two sources of uncertainty separately allowed comparison of the results from application of the linear equations for OLS.

After getting the distributions for the coefficients and each component of uncertainty, I used the 5th percentile and 95th percentile values to give me the range of values (range of coefficients and range of errors) for a 90% confidence level. Subtracting the 5th percentile value from the 95th percentile value, then dividing by 2 provided an interval for comparison with the OLS standard errors times the t-statistic.

Since this process was new to me, I used a large number of draws, from 10,000 to 100,000, while testing the process. I used a couple of different visualizations to make sure that I had sufficient draws, and that the results made sense. The most important were the histograms of the errors. An example, for 30,000 draws, $n=365$, and $m=180$, is shown in Figures 3-1 and 3-2. The highlighted bars are those nearest the 5th and 95th percentiles.

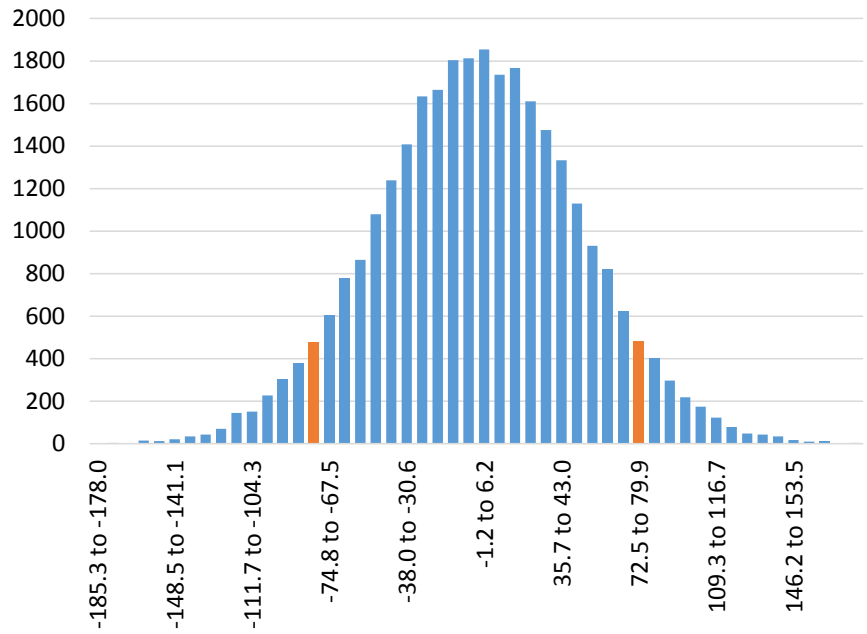


Figure 3-1. Histogram of Model Errors for 30,000 Bootstrap Samples

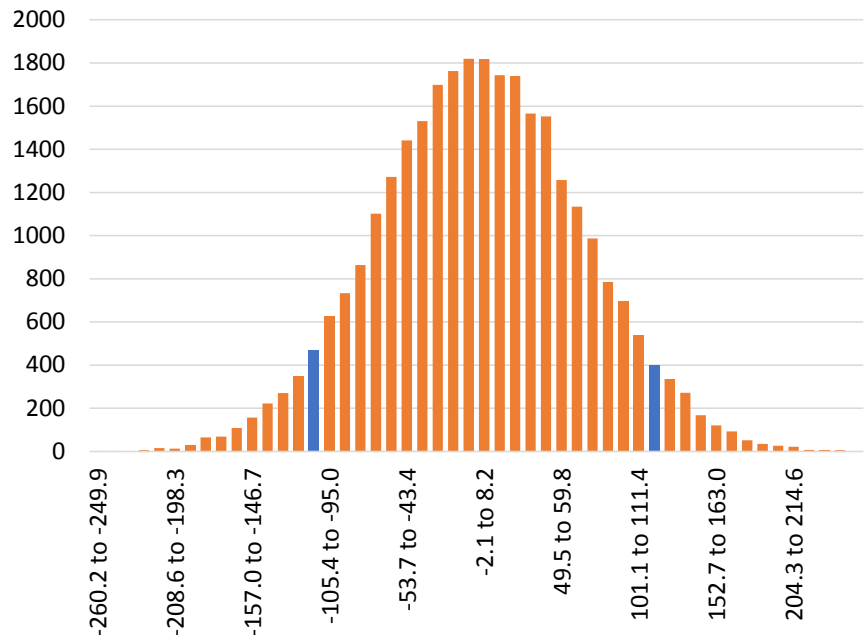


Figure 3-2. Histogram of Noise Errors for 30,000 Bootstrap Samples

The bins nearest the 5th and 95th percentile are highlighted. You can see that the histograms are quite smooth, indicating that the number of samples is sufficient.

3.4.1. Bootstrap Approach Assuming All Values are Independent

This approach was used for the simplest bootstrap analyses.

3.4.1.1. Data Set

The data set is summarized as:

- Synthetic data
- 365 data points
- No autocorrelation
- ~Normal distribution of residuals

A synthetic data set was used so that its characteristics would be known. It was based on a known linear relationship. The use of this synthetic data with known characteristics facilitated the initial comparison with the exact solution for the aggregate uncertainty for OLS. The use of 365 points was arbitrary, but that number obviously represents daily data.

The data, with the regression line, equation, and prediction limits for a 90% confidence level, are shown in Figure 3-3.

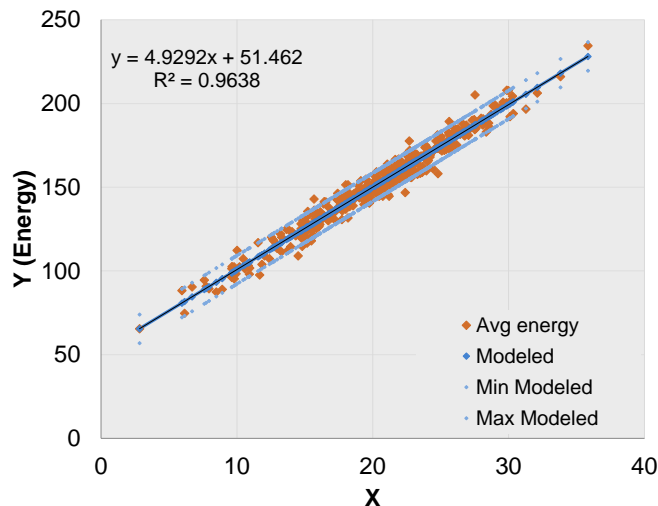


Figure 3-3. Regression Line, Equation, And Prediction Limits for the Synthetic Data

Figure 3-4 shows a lag chart. This chart plots, for the OLS model, the residuals at time $t+1$ versus the residuals at time t . Since there is no relationship, there is no autocorrelation. This is further confirmed by the Durbin-Watson statistic of 2.000. Durbin-Watson values vary between 0 and 4, with values significantly less than 2 indicating positive autocorrelation, and values significantly greater than 2 indicating negative autocorrelation. A value of 2 indicates no autocorrelation.

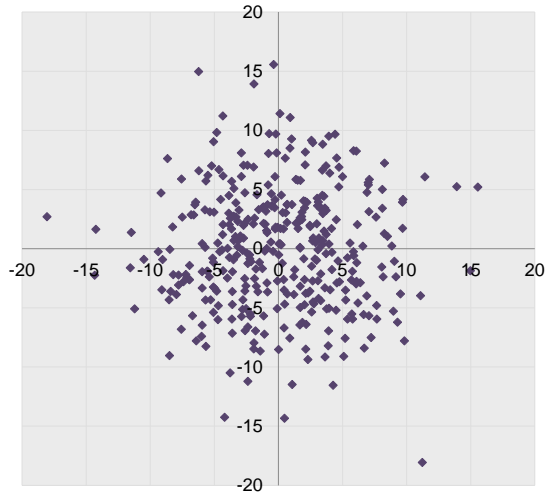


Figure 3-4. Lag Plot for Model Residuals for Synthetic Data Set

Figure 3-5 shows a histogram of residuals, which roughly approximates a normal distribution.

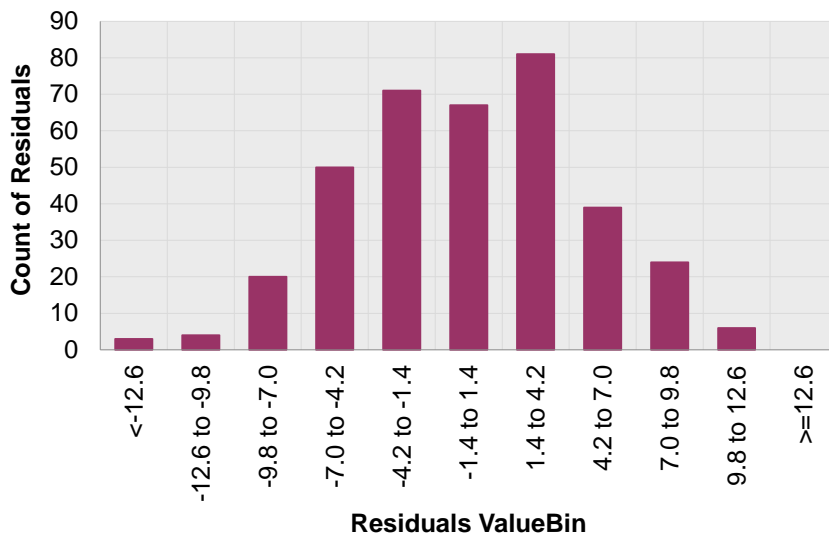


Figure 3-5. Histogram of Residuals for Synthetic Data Set

3.4.1.2. Bootstrap Process

Figure 3-6 shows the part of the spreadsheet with the bootstrap sampling.

iSamples												
30,000												
Intercept			Slope			mPost			Run Bootstrap			
51.183331			4.957609			180						
Totals:		55,697			55,096	55,204	108			53.7	0.9	54.6
M	N	O	P	Q	R	S	T	U	V	W	X	
Rand	Avg prod	Avg energy	Record number	Avg prod	Avg energy	Moded	Model Error	Rand2	Model Error	Model Residual	Total Error	
111	18.295115	138.504	1	15.21896	132.9941	126.633	0.154	196	0.337026	-8.46424	-8.12722	
95	12.621476	118.88	2	15.52848	124.2796	128.1675	0.163	23	0.320673	-1.71675	-1.39608	
202	26.242115	184.7821	3	22.49536	159.1026	162.7065	0.361	220	0.380398	-1.66925	-1.28885	
302	23.217608	169.7278	4	16.99809	131.6905	135.4532	0.205	66	0.337389	-3.98408	-3.64669	
187	16.119348	131.0174	5	19.007	143.6533	145.4126	0.262	246	0.365826	-2.96538	-2.59956	
229	25.579869	182.5789	6	8.47089	87.42839	93.1787	-0.037	363	0.258527	-0.24163	0.016897	
306	22.679374	177.5888	7	23.87815	175.1003	169.5619	0.400	126	0.321126	4.14255	4.463676	
289	28.122693	190.8132	8	27.9802	188.942	189.8982	0.517	114	0.338637	1.646535	1.985171	
113	15.671965	142.9598	9	20.09503	157.2123	150.8067	0.293	282	0.351705	4.539678	4.891383	
315	17.33219	142.8871	10	23.01077	170.544	165.2617	0.376	247	0.372042	1.074525	1.446567	

Figure 3-6. Portion of Spreadsheet Used for Bootstrap Resampling

The columns shown are Columns M through X in the workbook, as indicated by the letters above the colored area.

Column M provides a random number to draw the corresponding Record number—X,Y pair, Avg prod and Avg energy—in the bootstrap sample. The Rand of 111 indicates that the 111th X,Y pair is drawn, and their values are 18.29 and 138.50. Again, the draw can repeat data pairs, or omit data pairs entirely. In this particular draw, Record Number 1 was drawn 2 times, Record Number 2 was drawn 1 time, and Record Number 3 was not drawn at all. Record number 62 was drawn 5 times.

Column N has the X value associated with the record number from Column M.

Column O has the Y value associated with the record number from Column M

Using the set of drawn points represented in Columns N and O, OLS is used to get the Intercept and Slope coefficients shown above those columns. Farther up in Column N, below “iSamples,” is the number of bootstrap samples that the code will run.

Column P has the list of record numbers in the order of the original data.

Column Q has the X value associated with the record number from Column P.

Column R has the Y value associated with the record number from Column P.

In other words, Columns Q and R have the X and Y values for the original data, in the order of the original data. The spreadsheet has the original data in columns to the left of what is shown, but they were repeated in these columns for clarity.

Column S has the estimated values \hat{Y} for the Avg prod (X) value shown in column Q, based on the Intercept and Slope shown, which was estimated for this bootstrap sample.

Column T has the Model Error: the difference between the actual Y and \hat{Y} .

Columns M through T have 365 values, since there were 365 X,Y pairs in the original data.

Column U has the a random number to draw the additional bootstrap sample associated with this approach.

To clarify the two separate draws, they were used to separate the mean response (model) uncertainty from the noise uncertainty. The first sample was used to get the model coefficients. The second draw was to get the range of model residuals for the estimated coefficients. In hindsight, it may have been more computationally efficient to do a loop-in-loop set of draws—making multiple draws for each set of model coefficients—and fewer total model draws, but I made the second draw every time I made the first draw.

The number of points in the second bootstrap match the input number of points in the post (reporting) period. This value is shown above in Column U, under “mPost.” Therefore, this analysis was for 180 days in the reporting period, and Columns U through X each have 180 values in them.

Column V has the Model Error value from the record in column T for the random draw in column U. In other words, the value 0.337026 is for the 196th value in Column T.

Column W has the Model Residual from the original data (that residual is to the left of the data shown, in Column F) for the record number shown in Column U.

Column X adds the results in Column V and Column W to get the Total Error for each point.

The values above the letters “V”, “W”, and “X” indicating those columns are the totals for those columns. Each time a new pair of bootstrap samples is drawn (Columns 1 and 9), the associated Intercept and Slope, and the total Model Error, Model Residual (Noise), and Total Error are saved in columns Z through AD, as shown in Figure 3-7.

Y	Z	AA	AB	AC	AD
	Intercept	Slope	Model Error	Model Residual	Total Error
	51.27104	4.936832	(5.5)	42.0	36.5
	51.54564	4.920138	(18.0)	33.7	15.7
	51.06814	4.964354	54.1	(32.1)	22.1
	52.38379	4.920265	133.9	(41.9)	92.0
	51.66731	4.917121	(6.0)	(30.2)	(36.3)
	51.94632	4.918226	48.1	(15.7)	32.4
	51.77716	4.904323	(34.4)	10.8	(23.6)
	49.89925	4.995203	(45.5)	(43.9)	(89.3)
	52.37292	4.879445	(22.6)	116.4	93.8
	52.62343	4.891738	74.4	(10.9)	63.5

Figure 3-7. Bootstrap Samples’ Model Coefficients and Errors

There are 30,000 records in Columns Z through AD, corresponding to the value for iSamples in Column N.

These records are summarized above in Columns Y through AD, with an additional check in Column AE. Various percentile values Intercept and Slope, and the total Model Error, Model Residual, and Total Error, are calculated as indicated by Figure 3-8.

percentile	Intercept	Slope	Model Error	Model Residual	Total Error	QuadError
0.01%	47.4	4.75	(178.7)	(246.3)	(300.2)	304.3
5.00%	49.7	4.84	(79.3)	(111.4)	(137.0)	136.78
10.00%	50.1	4.86	(61.1)	(87.1)	(106.4)	106.4
20.00%	50.6	4.89	(40.2)	(57.6)	(71.5)	70.3
50.00%	51.5	4.93	(0.2)	(0.7)	(0.3)	0.7
80.00%	52.4	4.97	39.7	57.9	70.1	70.2
90.00%	52.8	4.99	61.1	87.8	106.8	106.9
95.00%	53.2	5.01	79.6	113.0	138.2	138.20
99.99%	55.1	5.12	182.2	255.9	317.4	314.1
	1.75	0.0838	79.46	112.20	137.59	137.49
Y	Z	AA	AB	AC	AD	AE

Figure 3-8. Percentiles of Bootstrapped Model Coefficients and Errors

The difference between the 95th and 5th percentile values are shown in the numbers at the bottom. These are the 90% confidence intervals estimated by the bootstrap. (In Bayesian terms, this is called a “credible interval.”)

Column AE adds a check on the Total Error, using quadrature to combine the Model Error and Model Residual. This check is not used for all the bootstrap approaches implemented, since it is not valid for some models and data sets, but it is valid here. As can be seen, that check provides essentially the same results as the Total Error column, except for slight differences at the tails of the distribution.

Columns AE through AL provide the data for the histograms of Model Errors and Model Residuals. The histograms use 50 bins, so there are 50 records in these columns. Figure 3-9 shows the first 10 records for the data for the histograms.

AE	AF	AG	AH	AI	AJ	AK	AL
Min of Bin	Max of Bin	Bin	Model Error Count in Bin	Min of Bin	Max of Bin	Bin	Model Residual Count in Bin
-178.7	-171.5	-178.7 to -171.5	3	-246.3	-236.3	-246.3 to -236.3	5
-171.5	-164.3	-171.5 to -164.3	5	-236.3	-226.2	-236.3 to -226.2	6
-164.3	-157.1	-164.3 to -157.1	6	-226.2	-216.2	-226.2 to -216.2	7
-157.1	-149.8	-157.1 to -149.8	11	-216.2	-206.1	-216.2 to -206.1	17
-149.8	-142.6	-149.8 to -142.6	11	-206.1	-196.1	-206.1 to -196.1	18
-142.6	-135.4	-142.6 to -135.4	26	-196.1	-186.0	-196.1 to -186.0	30
-135.4	-128.2	-135.4 to -128.2	45	-186.0	-176.0	-186.0 to -176.0	54
-128.2	-121.0	-128.2 to -121.0	64	-176.0	-166.0	-176.0 to -166.0	71
-121.0	-113.8	-121.0 to -113.8	84	-166.0	-155.9	-166.0 to -155.9	101
-113.8	-106.5	-113.8 to -106.5	125	-155.9	-145.9	-155.9 to -145.9	141

Figure 3-9. Data for Histograms of Residuals

This example used 30,000 bootstrap samples. Compare the histograms in Figures 3-9 and 3-10 with Figures 3-1 and 3-2. Note how the distributions became much smoother with the larger number of samples. Because I didn't want an insufficient number of samples to affect the results, 30,000 samples were used for this analysis.

3.4.2. Block Bootstrap Approach for Autocorrelated Residuals

A regular bootstrap does not retain the order of the data. 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 two data sets. First, a new synthetic data set was created. It used the same X-values as the prior synthetic data set, and was based on the same relationship, but it was created to have mild autocorrelation.

Figure 3-10 shows the lag plot for this data set. Compare with Figure 3-1, and note that this chart shows a slight positive autocorrelation. Correspondingly, the Durbin-Watson statistic for this data set is 0.93.

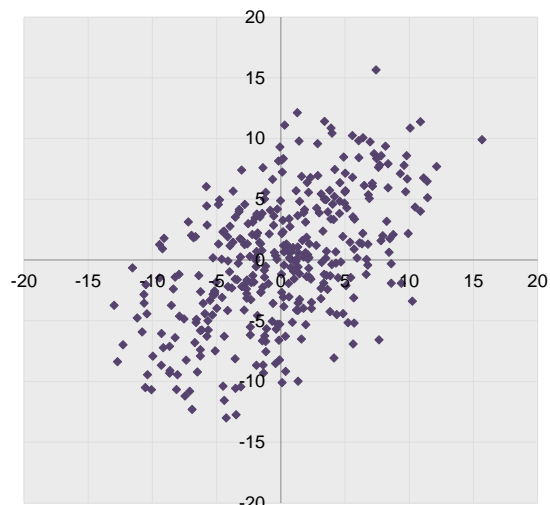


Figure 3-10. Lag Plot for Synthetic Data With Medium Autocorrelation

A key question with autocorrelation and a block bootstrap is, how big a block to use? E.g., for how many days is the energy use related? It depends on how many lags of autocorrelation are present. Is it only related to the prior day, i.e. lag 1? Is it related to the prior 4 days? One way to do this is with an autocorrelation plot, which indicates the relationship between points with different lags. I created some of these, but it was also intuitive to just try different block sizes and plot the results to show when the estimated aggregated error ceased to increase with increasing block size. Analyses for this first data set used block sizes of 1, 2, 4, 7, and 14 points.

Using a block size of 1 is the same as using a regular bootstrap, since the same number of points are in a block. As block sizes increase, the aggregated error was expected to increase since it would now include the effect of autocorrelation. Once the block size reaches the maximum lag present, the aggregated error would no longer change.

The second data set used for these analyses was again synthetic. However, the data set was created to have more scatter and more autocorrelation. Figure 3-11 shows the data set with the OLS fit line and the 90% prediction interval.

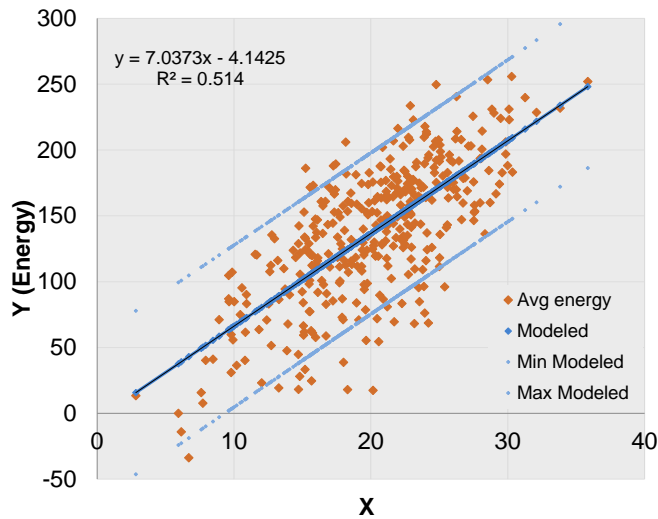


Figure 3-11. Data, Regression Line, and 90% Prediction Intervals for Synthetic Data with More Scatter and Significant Autocorrelation

Figure 3-12 shows the autocorrelation plot for this data set. The Durbin-Watson statistic is 0.61.

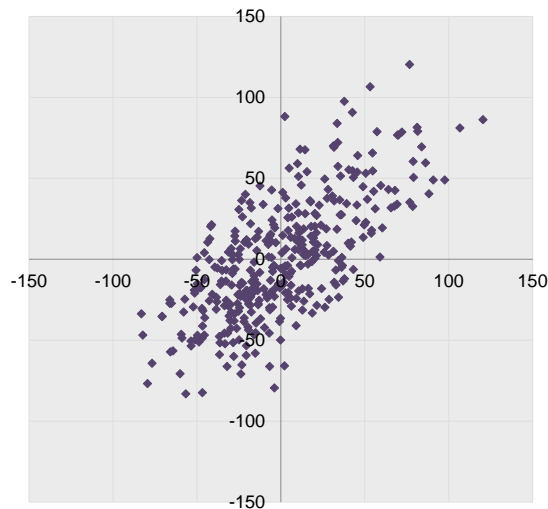


Figure 3-12. Lag Plot for Synthetic Data with Higher Autocorrelation

3.4.3. Bootstrap Approach for Data With a Relationship Between Independent Variable Values

The approach in the prior section was the foundation for the next two bootstrap analyses.

Another modified approach may be needed for common commercial and residential building energy models. Even with monthly data, which can reasonably be assumed to not have serial correlation, the data points are not independent. Since the X-axis independent variable is outside air temperature, there is a certain distribution of these temperatures that will occur over the course of a year. Ignoring this constraint on the range of independent variables could

affect the estimated uncertainty relative to the true uncertainty. The X-values are related because of the limitation on the range of temperatures likely at a given time of the year.

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.

The issue with this approach, I think, is that it assumes homoscedasticity—the residuals to be resampled can come from any part of the original regression, since they are randomly sampled, and applied to any part of the new data set being created.

In contrast, the approach of resampling data pairs, presented in the prior section, assumes independence of the X-values, but does not assume homoscedasticity.

To look at the impact of the X-values having a relationship with each other, three bootstrap analyses were performed:

- 0.0.
 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 X-values. With 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.

For this analysis, a real set of monthly data was used. This data set had a distinct change in slope. Hence, the least squares fit was for a 4-parameter model with a change point. The data and model are shown in Figure 3-13. No scales are shown since this is customer data.

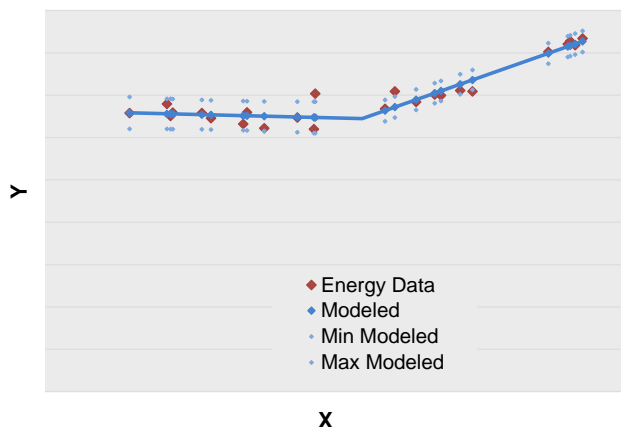


Figure 3-13. Monthly Data With a Relationship Between Independent Variable Values

This data set had no autocorrelation, since it was monthly data.

Note, however, that the prediction intervals show that there is heteroscedasticity. This affects the accuracy of the approaches, including the bootstrapping of the raw residuals. The

bootstrapping of the normal residuals is expected to provide the most accurate estimation of uncertainty.

Monthly data was chosen for several reasons:

- A poor distribution of temperatures is more likely to occur with monthly data, so the difference between resampling data pairs or residuals would be more apparent.
- There was interest in finding what a bootstrap approach would find for the most likely change point, since it was somewhat ambiguous for this data set.
- It is faster to process monthly data because of the relatively few number of data points.

The process changes for this data set and resampling of residuals, relative to the approach for the synthetic data set that met all the requirements for OLS regression, were as follows:

X-values (temperatures) were not resampled.

Y-values (energy) were calculated as the OLS fit for that temperature and adjusted based on either a resampled residual, or a random value from a normal distribution of residuals.

The coefficients to be estimated included four parameters instead of two:

- 0.0.
 1. Intercept
 2. Left Slope
 3. Change Point temperature
 4. Right Slope

For each bootstrap sample, the additional coefficients were saved along with the Model Error, Model Residual, and Total Error.

In the remaining aspects, the process was the same as the process described for the synthetic data set without autocorrelation.

4. RESULTS FOR EACH METHOD ANALYZED

There were four data sets analyzed. These were described in the preceding section, and are summarized here:

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, 4-parameter relationship, X-values not independent

A point of clarification in the results shown in this section: The “errors” quantified are not a bias—all of the models were unbiased. Instead, the errors are one side of a confidence band on a savings estimate.

4.1. Results for Synthetic Data with a Linear Relationship and No Autocorrelation

The coefficients and uncertainties were estimated for these approaches”

- OLS
- Bootstraps
- ASHRAE Fractional Savings Uncertainty
- Improved ASHRAE Fractional Savings Uncertainty

The uncertainties were estimated for the following lengths of reporting periods (days):

- 30
- 60
- 90
- 180
- 365

The bootstraps for these analyses used 10,000 samples. This appeared to be near the minimum to get smooth histograms of errors. Figures 4-1 and 4-2 show the histograms of Model Errors and of Residuals Errors respectively

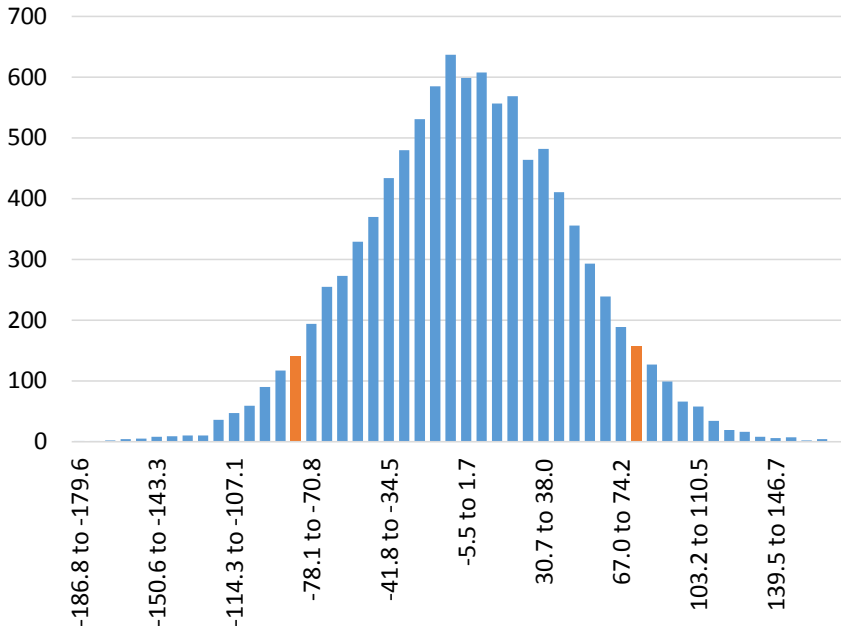


Figure 4-1. Histogram of Model Errors for 10,000 Bootstrap Samples

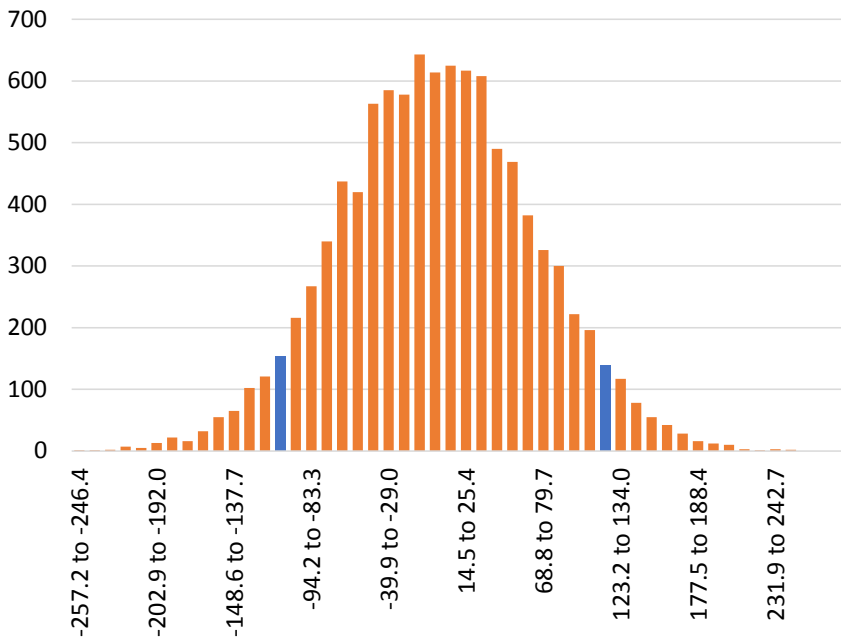


Figure 4-2. Histogram of Noise Errors for 10,000 Bootstrap Samples

The histogram in Figure 4-2 is a little rough, indicating that the number of samples is near a minimum for reliable results. Contrast these figures with Figures 3-1 and 3-2 and the impact of the decreased number of samples can be seen.

4.1.1. Coefficients

Table 4-1 summarizes the results from the estimation of coefficients and their errors. The first 5 columns show the 50th percentile value from the bootstraps and the last column is for OLS.

Table 4-1. Comparison of Coefficients and their Errors for Bootstrap and OLS Approaches

mPost	30	60	90	180	365	OLS
Intercept at X=0	51.461	51.456	51.456	51.457	51.467	51.462
Slope	4.92967	4.92861	4.92933	4.92931	4.92901	4.92919
Error of the Intercept	1.762	1.797	1.792	1.735	1.757	1.725
Error of the Slope	0.08440	0.08489	0.08497	0.08342	0.08311	0.08263

Note: For the coefficient errors for OLS, the Error is the t-statistic (1.649) X the Standard Error

The results are very close for all parameters. This is expected, because the bootstrap uses OLS to estimate the coefficients for each sample drawn, and hence the 50th percentile values should be very close to the corresponding OLS coefficients

4.1.2. Uncertainty

Table 4-2 summarizes the equations used for the formulaic (non-bootstrap) estimations of uncertainty, summarizing Section 3. For definitions of variables, refer to Section 3. Note that the ASHRAE approaches do not divide the uncertainty into parts, but only estimate total uncertainty.

In Tables 4-2 and 4-3, “aggregate error” means the error associated with adjusted baseline predictions for all the points, i.e. the aggregate for all of the time intervals.

Table 4-2. Uncertainty Equations

Model aggregate error by OLS	$t \times se(\hat{y}) = t \times \frac{s \times m_{post}}{\sqrt{n_{base}}} \times \left(1 + \frac{(\bar{x}_{base} - \bar{x}_{post})^2}{\text{Var}(x_{base})} \right)^{\frac{1}{2}}$
Noise aggregate error by OLS, with adjustment for autocorrelation	$t \times s \times \sqrt{m_{post} \times n_{base} / n_{prime}}$
Total aggregate error by OLS	$\sqrt{(\text{Model aggregate error})^2 + (\text{Noise aggregate error})^2}$
Total aggregate ASHRAE FSU	$\frac{\Delta E_{save,m}}{E_{save,m}} = t \times \frac{1.26 \cdot CV \left[\frac{n}{n'} \left(1 + \frac{2}{n'} \right) \frac{1}{m} \right]^{1/2}}{F}$
Total aggregate Improved ASHRAE FSU	Data Sets 1 through 3, daily data: $\frac{\Delta E_{save,m}}{E_{save,m}} = t \times \frac{(aM^2 + bM + c) \cdot CV \left[\frac{n}{n'} \left(1 + \frac{2}{n'} \right) \frac{1}{m} \right]^{1/2}}{F}$

Table 4-3 summarizes the results from the estimation of model and noise errors for OLS and bootstraps, and total errors for all four methods. Note that the ASHRAE approaches do not divide the uncertainty into parts, but only estimate total uncertainty.

These values are for an 90% confidence level.

Table 4-3. Comparison of Savings Errors for All Four Approaches for Data Set 1

mpost	30	60	90	180	365
Model aggregate error by OLS	13.2	26.5	39.7	79.4	160.9
Model aggregate error by Bootstrap	13.5	27.0	40.2	79.4	158.8
Noise aggregate error by OLS	46.1	65.2	79.9	113.0	160.9
Noise aggregate error by Bootstrap	45.9	64.5	79.0	112.0	157.0
Total aggregate error by OLS	48.0	70.4	89.2	138.1	227.6
Total aggregate error by Bootstrap	48.2	69.7	89.1	135.9	225.5
Total aggregate error by ASHRAE FSU	58.3	82.4	101.0	142.8	203.3
Total aggregate error by Improved FSU	48.0	70.1	88.6	136.4	224.7

Again, note that all of the values are quite close, except the original ASHRAE FSU approach. Figure 4-3 shows the model and noise errors in chart form. Figure 4-4 shows the total errors for each approach.

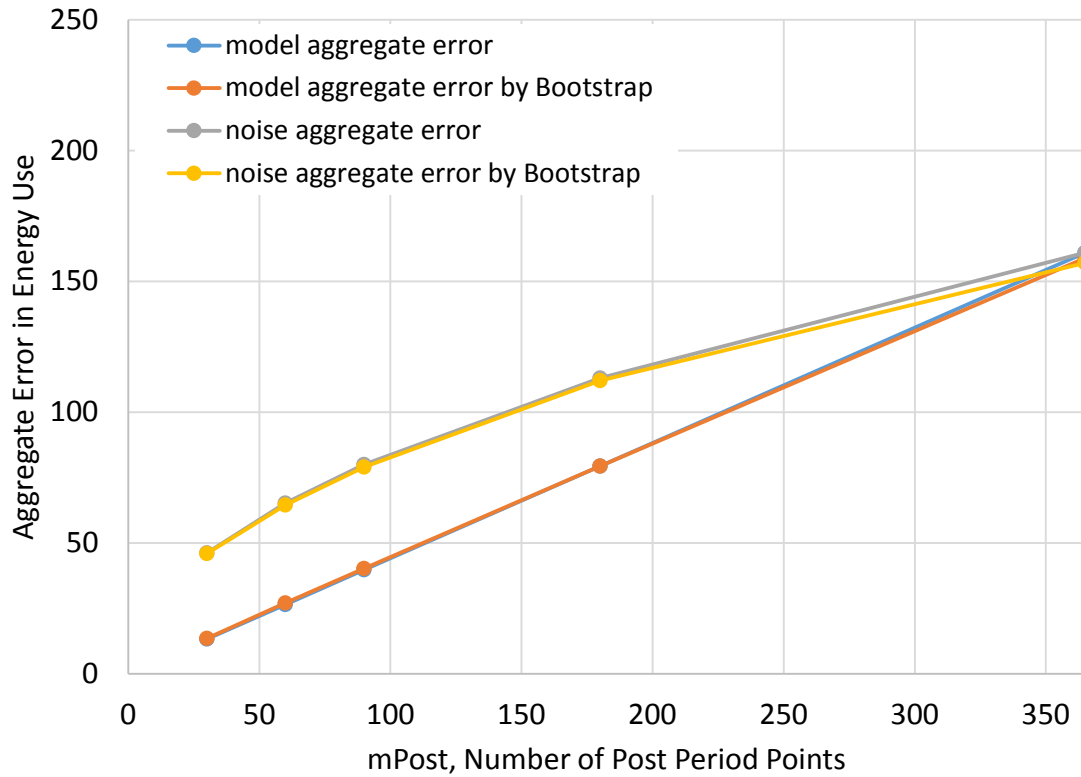


Figure 4-3. Model and Noise Aggregate Errors by Approach and Length of Reporting Period

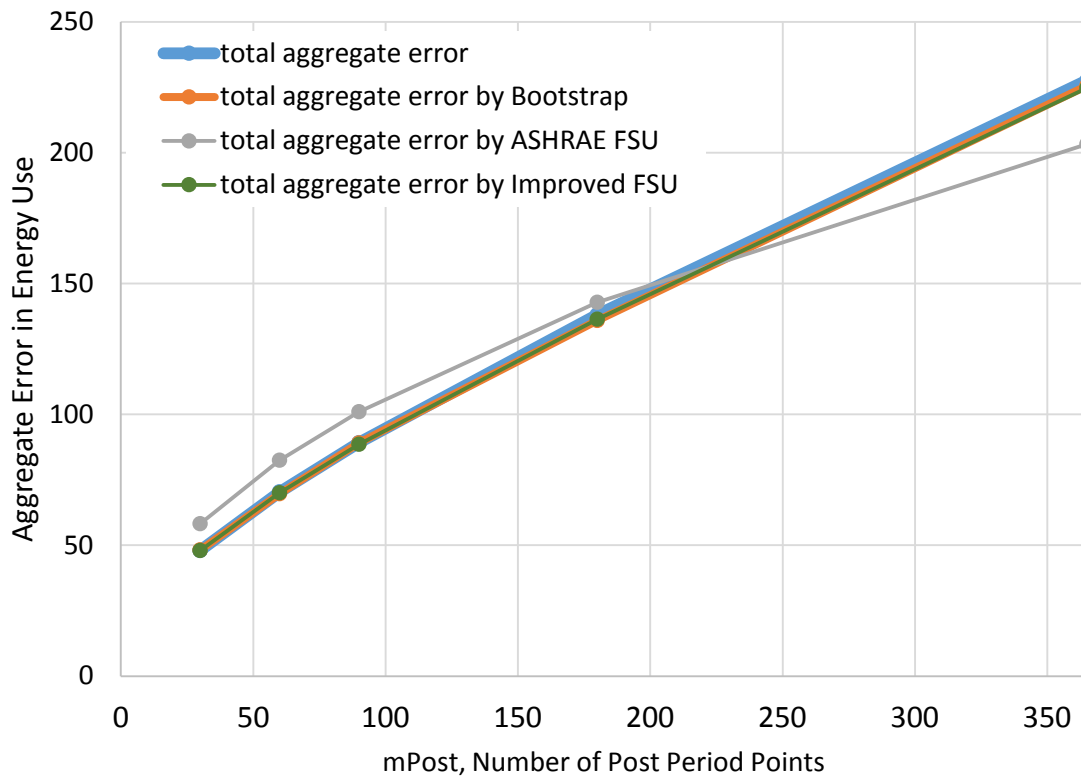


Figure 4-4. Total Aggregate Errors by Approach and Length of Reporting Period

4.2. Results for Synthetic Data with a Linear Relationship and Moderate Autocorrelation

These analyses were only performed for reporting periods of one year.

To estimate the block sizes needed to account for the autocorrelation, an autocorrelation plot (also called a correlogram) was created. Figure 4-4 shows the autocorrelation plot.

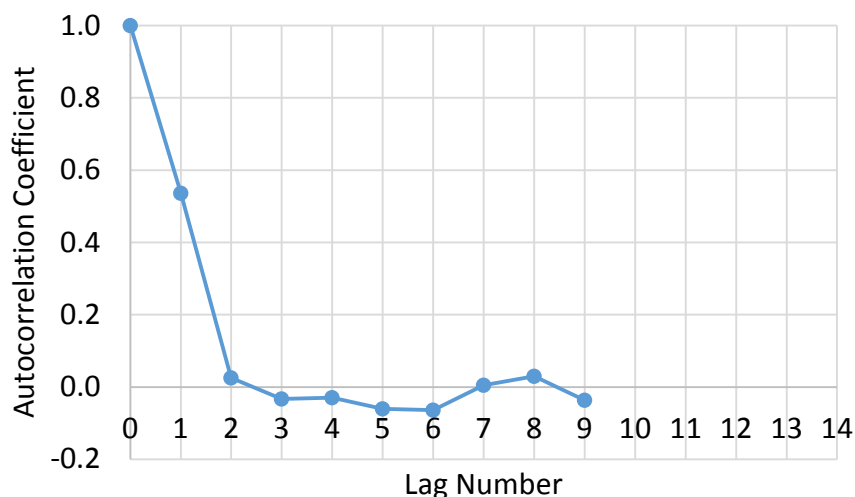


Figure 4-5. Autocorrelation Plot for Data Set 2

Figure 4-5 shows that only the Lag-1 Autocorrelation Coefficient is significant. Hence, a block size of two might be all that is required to fully account for autocorrelation using a block bootstrap. Despite this inference, I ran the bootstrap for all the block sizes listed above. Also, only the basic bootstrap without blocks was run for only 10,000 samples. The other analyses were run for 30,000 bootstrap samples, with checks using 100,000 samples. Table 4-4 shows the results of the analyses for Data Set 2.

Table 4-4. Comparison of Savings Errors for All Four Approaches for Data Set 2.

OLS: No Autocorrelation			+ ASHRAE Autocorr. Adjust.			ASHRAE FSU		Block Bootstrap				
Model Error	Noise Error	OLS Total	Model Error	Noise Error	OLS + Autocorr.	FSU	Improved FSU	Block Size	Samples	Model Error	Noise Error	Total Error
169	169	239	308	308	435	392	433	1	10,000	169	169	240
169	169	239	308	308	435	392	433	2	30,000	210	209	295
169	169	239	308	308	435	392	433	4	30,000	226	227	319
169	169	239	308	308	435	392	433	4	100,000	228	226	319
169	169	239	308	308	435	392	433	7	30,000	227	230	324
169	169	239	308	308	435	392	433	7	100,000	230	229	326
169	169	239	308	308	435	392	433	14	30,000	229	228	322
169	169	239	308	308	435	392	433	14	100,000	228	229	323
84	119	146	84	119	146	276		7	30,000	229	229	324

This information may be seen more clearly in Figure 4-6.

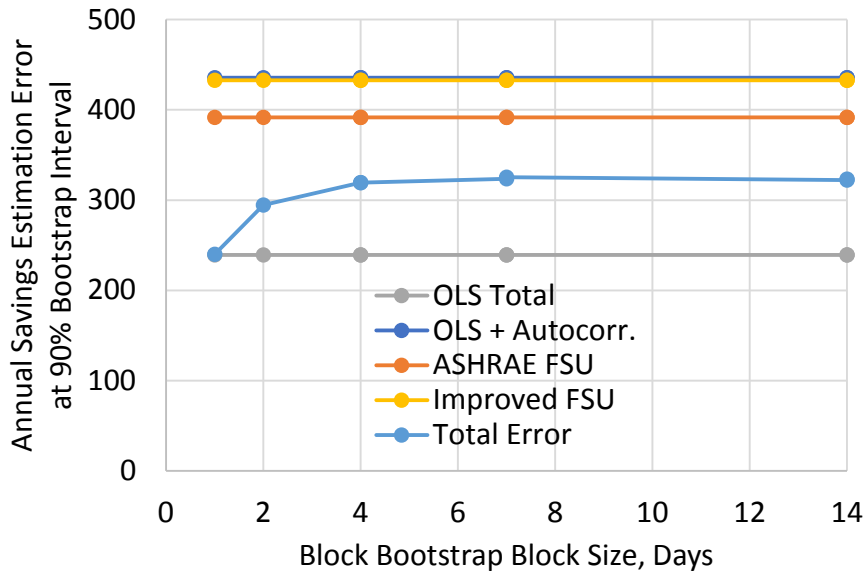


Figure 4-6. Total Error vs. Block Size for Data Set 2

Figure 4-6 shows that a block size of 4 is needed before the estimated error ceases to increase. This is higher than expected based on the autocorrelation plot.

More importantly, it shows, as expected, that not accounting for autocorrelation underestimates the error, as shown by the values for OLS and the bootstrap without blocks (i.e. block size =1.)

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 the same adjustment for effective number of points as the FSU equations. Since the OLS and improved FSU equations provided the same result for Data Set 1, which didn't have autocorrelation, it is not surprising that they provide the same result here since they use the same adjustment for autocorrelation.

However, even the improved FSU equation does not provide results that match the block bootstrap. It estimates an error at the 90% confidence level that is about 35% higher (433/319 – 1) than the bootstrap.

I theorize some reasons why the Improved FSU equation gives different results:

- The adjustment for the effective number of points uses only the Lag-1 autocorrelation coefficient, and higher order lags appear to be important.
- The adjustment for the effective number of points is itself an approximation.

These are just possibilities. It is also possible that the block bootstrap is not providing the correct results, possibly because of an incorrect implementation.

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

The same block sizes as above were used for these analyses, plus a block size of 5.

Figure 4-7 shows the autocorrelation plot for Data Set 3. It shows that lags up to 5 may be important.

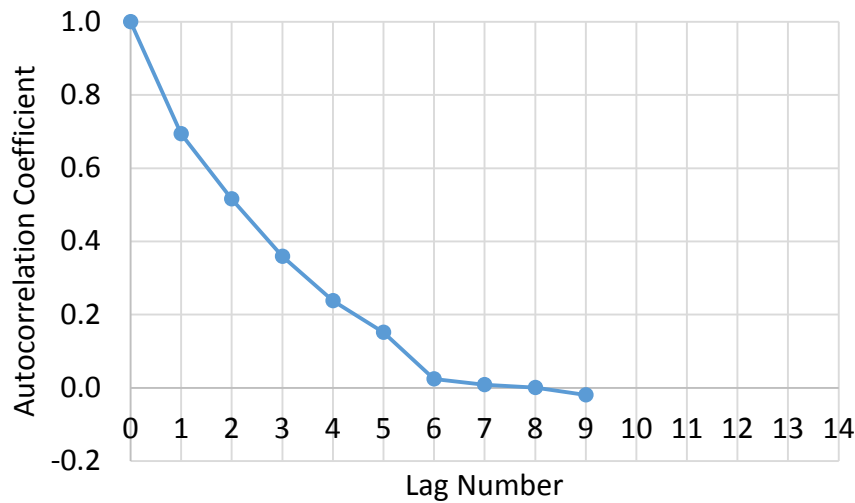


Figure 4-7. Autocorrelation Plot for Data Set 3

The bootstrap was run for block sizes of 1, 2, 4, 5, 7, and 14. The results are shown in Table 4-5.

Table 4-5. Comparison of Savings Errors for All Four Approaches for Data Set 3

OLS: No Autocorrelation			+ ASHRAE Autocorr. Adjust.			ASHRAE FSU		Block Bootstrap				
Model Error	Noise Error	OLS Total	Model Error	Noise Error	OLS + Autocorr.	FSU	Improved FSU	Block Size	Samples	Model Error	Noise Error	Total Error
1164	1166	1648	2747	2747	3885	3513	3883	1	10,000	1164	1166	1648
1164	1166	1648	2747	2747	3885	3513	3883	2	10,000	1505	1495	2118
1164	1166	1648	2747	2747	3885	3513	3883	4	30,000	1903	1903	2688
1164	1166	1648	2747	2747	3885	3513	3883	4	100,000	1909	1901	2690
1164	1166	1648	2747	2747	3885	3513	3883	5	100,000	2037	2028	2883
1164	1166	1648	2747	2747	3885	3513	3883	7	30,000	2187	2180	3090
1164	1166	1648	2747	2747	3885	3513	3883	14	30,000	2332	2327	3271

This information may be seen more clearly in Figure 4-8.

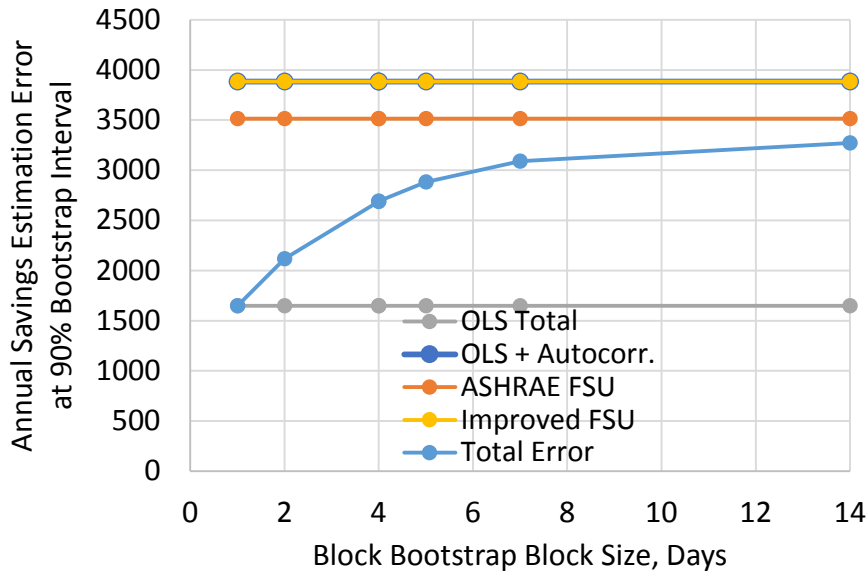


Figure 4-8. Total Error vs. Block Size for Data Set 3

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%.

Again, the uncertainty estimated by the improved FSU equation matches the OLS with the same adjustment for the effective number of points. Also, the FSU equations continue to estimate higher uncertainties than the block bootstrap.

4.4. Results for Real Data with a 4-Parameter Relationship, X-Values Not Independent

Table 4-6 shows the results for the OLS approach, both FSU approaches, and three resampling approaches. These analyses were for a 24-month baseline and a 24-month reporting period.

Table 4-6. Estimated Uncertainties by Approach for a Data Set With Related X-Values

OLS			ASHRAE FSU		Bootstrap Resampling Method								
					Data Pairs			Residuals			Normal Residuals		
Model Error	Noise Error	Total Error	Improved FSU	FSU	Model Error	Noise Error	Total Error	Model Error	Noise Error	Total Error	Model Error	Noise Error	Total Error
104,677	104,677	148,036	137,279	175,113	118,045	92,462	145,708	99,911	92,728	134,914	107,647	90,895	139,212

The OLS approach might not be considered applicable, since the model is not linear. However, its utility for these applications is of interest. OLS showed the highest uncertainty. This is not surprising, since OLS assumes that all data pairs are independent, but here the X-values are not independent. Surprisingly, the improved FSU shows much higher uncertainty than the OLS approach.

Of the bootstrap methods, the resampling of data pairs had the highest uncertainty. This was expected, since samples could have mostly cold temperatures, or mostly high temperatures. Because the data should actually be constrained to reasonable distributions of temperatures, and because of heteroscedasticity in the residuals, a relatively high estimate of uncertainty was expected.

The resampling of residuals showed lower uncertainty than the resampling of normal residuals. This is surprising, because the heteroscedasticity should mean a wider distribution of sampled residuals. I suspect that the implementation of the normal residuals could be at fault here. The implementation 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 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.

Figure 4-10 shows these results for the total errors.

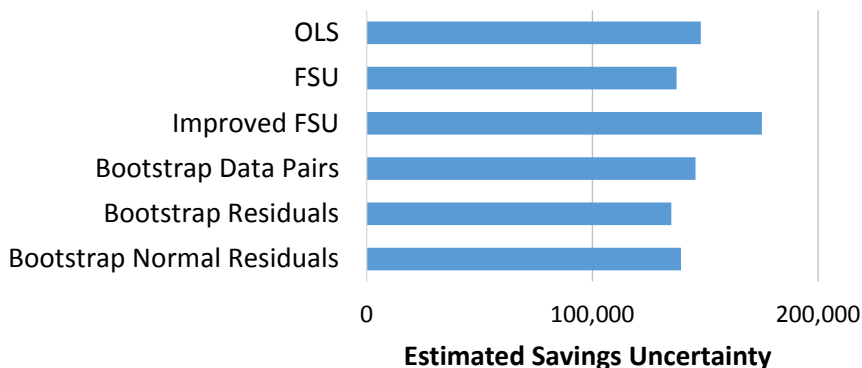


Figure 4-10 Estimated Uncertainties by Approach for a Data Set With Related X-Values

For this data set the improved calculation did not make a large difference

5. CHOSEN APPROACH FOR UNCERTAINTY ESTIMATION IN ECAM

ECAM version 5 implements both the Rushton equations for OLS and the Improved FSU equation. The OLS equations are modified in a manner similar the ASHRAE approach to handling autocorrelation. However, the OLS equations apply the ASHRAE autocorrelation adjustment only to the noise portion of the uncertainty, since this appears to be more accurate for these data sets.

These approaches meet the current need for a reasonably accurate estimate of uncertainty while not requiring excessive computation time.

The OLS equation is used for the actual estimation of uncertainty in ECAM and the Improved FSU equation is included as a reference to a known standard. Since it is based on the two separate components of uncertainty—model uncertainty and noise—as described in Section 2, I trust the OLS equations more than the ASHRAE equations to still provide reasonable results when applied to situations that don't fit the requirements for linear regression. Users should still be careful in trusting results for data with significant autocorrelation.

6. 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 more 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.