Table of Contents for Document (DRAFT)

[Introduction 1](#_Toc407793731)

[Sources of Uncertainty 1](#_Toc407793732)

[Systematic Error 1](#_Toc407793733)

[Random Error 2](#_Toc407793734)

[Definitions of Statistical Terms 3](#_Toc407793735)

[Sample Size Determination 8](#_Toc407793736)

[Determination of Metering Period 10](#_Toc407793737)

[Modeling 10](#_Toc407793738)

[Coefficient of Determination (R2) 11](#_Toc407793739)

[t-statistic 12](#_Toc407793740)

[Modeling Errors 13](#_Toc407793741)

[Omission of Relevant Variables 13](#_Toc407793742)

[Inclusion of Irrelevant Variables 14](#_Toc407793743)

[Functional Form 14](#_Toc407793744)

[Data Shortage 14](#_Toc407793745)

[Evaluating Regression Models 14](#_Toc407793746)

[Overfitting 16](#_Toc407793747)

[Combining Components of Uncertainty 16](#_Toc407793748)

# Introduction

A significant challenge to estimating savings that result from the installation of energy efficiency equipment is the lack of direct measurement. One can directly measure energy *consumption*, but not energy *savings,* which is the difference between actual consumption and what consumption *would have been* had energy efficiency measures not been installed.

Measurement errors, defined as the difference between estimates and actual values, are an expected outcome of any analytical method. Uncertainty can be introduced at every stage of the measurement and verification (M&V) process, including sampling, measurement, and adjustment. It is nearly impossible to quantify the effect of every potential source of error. M&V reports often limit uncertainty discussions to random error (especially sampling error and regression error). However, every effort should be made to identify and attempt to minimize every potential source of uncertainty.

The equation below illustrates the typical approach to estimating project savings. This approach involves potential measurement errors at the following stages: (1) estimating baseline energy use, (2) estimating the reporting period energy use, and 3) applying adjustments.

*Savings = (Baseline-Period Use - Reporting-Period Use) ± Adjustments*

An analytic method is efficient if it considerably reduces uncertainty relative to the study’s cost. There are many strategies available for designing an efficient study.

## Sources of Uncertainty

Uncertainty is an overall indicator of how well a calculated or measured value represents a true value. Without some measurement of uncertainty, it is impossible to judge an estimate’s value as a basis for decision-making.

Ideally, M&V efforts are designed to reliably determine energy and demand savings with some reasonable accuracy. However, this objective can be affected by either systematic error (i.e., not occurring by chance, including measurement error) or random error (i.e., occurring by chance and often due to using a sample rather than a census to develop the calculated or measured value).

## Systematic Error

Sources of systematic errors include the following three areas:

* **Data Measurement.** At this stage, errors are typically caused by meter reading errors, technicians incorrectly recording data, equipment failure, incorrect meter placement, or poor calibration. Measurement errors are best handled through using better metering equipment and improving data collection processes. In most applications, this error source is ignored. This is appropriate when using utility-grade electricity or natural gas metering equipment or if other metering equipment is of high-caliber. Human errors need to be tracked through quality control processes, which can then be quantified and included in the overall uncertainty assessment. For mechanical devices―such as meters or recorders―it is possible to perform tests with multiple meters to assess the measurement variability. However, it is more practical to use manufacturer or industry study information on the likely amount of error for any single piece of equipment for most of the data collection devices typically used in M&V studies,.
* **Data Collection**. Non-coverage errors can occur when some parts of a population are not included in the sample. This may include metering a sample of equipment that is not representative of the population of equipment. Non-coverage error is reduced by investing in a sampling plan that addresses known coverage issues. The M&V plan must clearly demonstrate that the sample is indeed representative by asking the following questions: *Was the sample frame carefully evaluated to determine what portions of the population, if any, were excluded in the sample? If so, what were steps taken to estimate the impact of excluding this portion of the population from the final results?* For example, the extrapolation to non-metered equipment must address the needed bias correction.
* **Data Modeling.** Estimates are calculated using models. Some models are fairly simple and straightforward (e.g., estimating the mean), and others are complicated (e.g., estimating response to temperature through regression models). Regardless, modeling errors may occur for a number of reasons such as using the wrong model, assuming inappropriate functional forms, including irrelevant information, or excluding relevant information. The M&V plan must illustrate the theoretical rational for the modeling approach and the inclusion of the relevant variables. When such variables are not available for modeling, their impact on the final estimates must be discussed and ways of reducing the introduced bias addressed. Also, the plan should address the process used for selecting formulas. Evaluators should also consider the following questions: *Are the models and adjustments conceptually justified? Have the sensitivity of estimates to key assumptions required by the models been tested or explained? Is the final model sensitive to inclusion or exclusion of explanatory variables? Is it sensitive to individual data points?*

Many M&V studies do not report any uncertainty measures besides a sampling error-based confidence interval for estimated energy or demand savings values. This is misleading because it suggests potentially incorrect assumptions: (1) the confidence interval describes the total of all uncertainty sources (which is incorrect), or (2) the other sources of uncertainty are not important relative to sampling error. Sometimes, however, uncertainty due to measurement and other systematic sources of error can be significant.

## Random Error

Any selected sample is only one of a large number of possible samples of the same size and design that could have been drawn from its population. Random errors are the result of using samples instead of the whole population. This may be the result of metering a sample of lighting fixtures or metering all lighting fixtures for a period of time (i.e., not sampling all the lights all the time).

Sampling creates errors because not all units under study are measured. The simplest sampling situation is that of randomly selecting *n* units from a total population of N units. In a random sample, each unit has the same probability of being included in the sample.

In general, the amount of error is inversely proportional to. That is, increasing the sample size by a factor “f” will reduce the error (improve the precision of the estimate) by a factor of.

To summarize, M&V plans need to discuss the expected uncertainty associated with:

1. Metering equipment. Discuss the metering equipment chosen. Have the meters been tested? What are the manufacturers claim for accuracy?
2. Modeling approach. Clearly explain the models to be used. What are the dependent and independent variables? Are there any important variable left out? Why? What the plan for conducting sensitivity analysis with independent variables and individual data points?
3. Sampling design. Clearly lay out the sample design. How my end uses will be metered? For how long? Potential impact of metered period relative to a full year? Expected levels of confidence and precision?

Before proceeding further, some statistical definitions are in order.

### Definitions of Statistical Terms

In general, statistics involves the following three types of data measurements:

* *Central Tendency.* This includes estimates of the mean, median, and mode. The mean is described below. The median is the middle point of a data set when arranged in ascending or descending order. The mode is the most frequently observed value.
* *Dispersion.* This includes measurements of the variance, the standard deviation, and the standard error.
* *Association.* This includes correlation and causation.

Descriptive estimators—such as the mean and standard deviation—can be calculated for any data set. The mean is the arithmetic average of the values, while the standard deviation is a measure of the variability among observations in the data. In normally distributed data, about 68% of observations are within one standard deviation of the mean, and 95% are within two standard deviations. (Note that a large standard deviation indicates greater dispersion of individual observations about the mean.)

The exact value of an estimate depends on the particular sample drawn. Thus, if an entire M&V process were repeated multiple times with a different sample drawn each time, a different estimated value would result for each process.

**Sample Mean ():** The sample mean is determined by adding up the individual data points (Yi) and dividing by the total number of these data points (n), as follows:



**Sample Variance (s2):** Sample variance measures the extent to which observed values differ from each other (i.e., variability or dispersion). The greater the variability, the greater the uncertainty in the mean. Sample variance is found by averaging the squares of the individual deviations from the mean. The reason these deviations from the mean are squared is simply to eliminate the negative values (when a value is below the mean) so they do not cancel out the positive values (when a value is above the mean). Sample variance is computed as follows:



**Sample Standard Deviation (s):** This is simply the square root of the sample variance. The sample standard deviation brings the variability measure back to the units of the data (as the computation above produces estimate of variation in squared units; e.g., the variance units are kWh2, the standard deviation units would be kWh).



**Sample Standard Error (SE):** This is the sample standard deviation divided by. This measure is used in estimating precision of a sample mean.



**Sample Standard Deviation of the Total (stot):** Many times evaluators are interested in the statistical properties of a total rather than a mean. The sample standard deviation of a total is used to define the precision about a sample total. It is defined as the square root of the sample size,  times the sample standard deviation:



**Coefficent of Variation (cv):** The coefficient of variation is simply the standard deviation of a distribution expressed as a percentage of the mean. This is equivalent to the inverse of the signal-to-noise ratio. The general formula is shown below.



**Precision:** Precision is the measure of the absolute or relative range within which the true value is expected to occur with some specified level of confidence. Confidence level refers to the probability that the quoted range contains the estimated parameter.

**Absolute Precision:** This is computed from sample standard error using a “t” value from a “t-distribution,” as shown in the formula below:

t  SEȲ

Table 1 shows at-distribution table and can also be found in statistic tables, books, or online resources.

Table . t-Distribution Table



Note: Calculate degrees of freedom (DF) using: *DF = n – 1 (for a simple computations like estimating the mean) or n – k – 1 (for a regression model)* Where n = sample size and k= number of explanatory variables.

In general, the true value of any statistical estimate is expected fall within a given confidence level; the range of the estimate is defined by the following formula:

*Range = estimate ± absolute precision*

Where “estimate” is any empirically derived value of a parameter of interest (e.g., total consumption, average number of uni8ts produced, etc.).

Relative precisionis the absolute precision divided by the estimate, as illustrated below:



### Data Analysis Example

The following example illustrates how to practically apply the statistical terms and formulas previously described. Consider the data in Table 2 from 12 monthly readings of a meter and the related analysis of the difference between each reading. The 12 values shown in the table may be considered a sample of one piece equipment being metered from many or the only piece equipment present, but it is a sample as in it was metered for one year to determine average use over many years.

Table . Data Analysis Example



The mean value is:



The variance is:



The standard deviation is:



The standard error is:



In Table 2 there are 12 data points. That means the DF= 12-1=11. Using Table 1 as a reference, for a confidence level of 90%, the value for “t” is 1.80.

Therefore, the Absolute Precision is:



And the Relative Precision is:



So, there is 90% confidence that the true mean-monthly consumption lies in the range between 923 and 1,077 kWh. It can be said with 90% confidence that the mean value of the 12 observations is 1,000 ±7.7%. Similarly, it could be said that with:

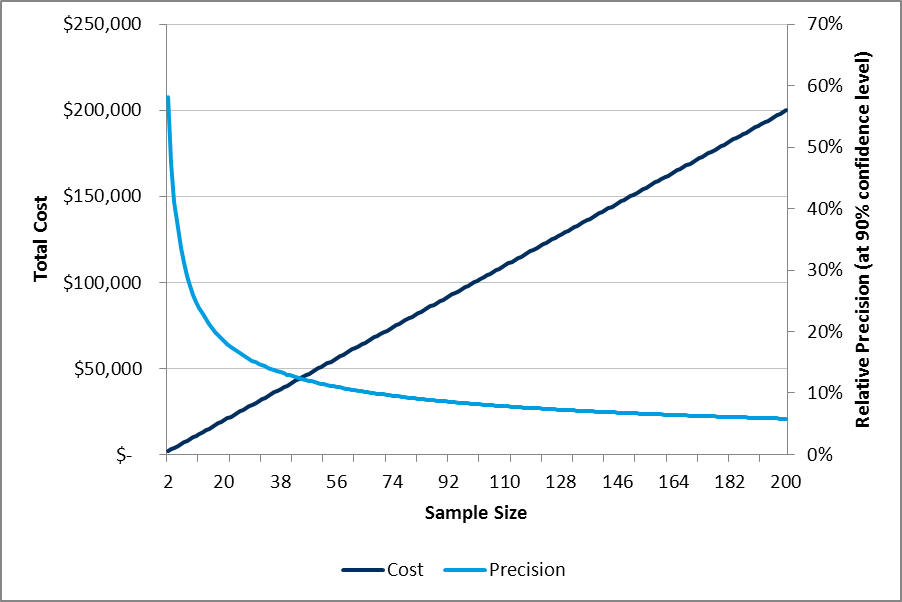
* 95% confidence, the mean value of the 12 observations is 1,000 ±9.5%;
* 80% confidence, the mean value of the 12 observations is 1,000 ±5.8%; or
* 50% confidence. the mean value of the 12 observations is 1,000 ±3.0%.

In general, high levels of confidence can be achieved with wider intervals, while narrower (more precise) intervals permit less confidence. In other words, there is a trade-off between precision and confidence. As a result, any statement of precision without a corresponding confidence level is incomplete.

There is always a trade-off between cost and precision. Increasing the sample size always leads to better precision. However, the cost of doing so can be prohibitive. The general precision equation can be written in this form:

The confidence level is fixed for a given study (typically at 90%). The population variance does not change with sample size either, so the only factor under the M&V control is the sample size. However, precision is not improved at a rate proportional to the sample size, but by the square root of the sample size. This is an important consideration in evaluation planning, as costs-per-sample-unit are often linear, while improvements in precision are not. Figure 1illustrates the trade-off for a hypothetical M&V process.

Figure . Example of Cost versus Precision



## Sample Size Determination

You can minimize sampling error by increasing the fraction of the population that is sampled. Increasing the sample size typically increases cost. As several issues are critical in optimizing sample sizes, the following steps should be used to set the sample size.

1. *Select a homogeneous populations, technologies, or end uses*. In order for sampling to be cost effective, the measured units should be expected to be the same as the entire population. If there are two different types of units in the population, they should be grouped and sampled separately. For example, when designing a sampling program to measure the operating periods of room lighting controlled by occupancy sensors, rooms occupied more or less continuously (e.g., multiple person offices) should be separately sampled from offices only occasionally occupied (e.g., meeting rooms). Clearly defining the target population being sampled is critical to the validity of the analysis.
2. *Determine the desired precision and confidence levels for the estimate (e.g., hours of use) to be reported.* Precision refers to the error bound around the true estimate (i.e., ±x% range around the estimate). Higher precision requires larger sample. Confidence refers to the probability that the estimate will fall in the range of precision (i.e., the probability that the estimate will indeed fall in the ±x% range defined by the precision statement). Higher probability also requires larger samples. For example, if you want 90% confidence and ±10% precision, you mean that the range defined for the estimate (±10%) will contain the true value for the whole group (which is not observed) with a probability of 90%. As an example, in estimating the lighting hours at a facility, it was decided to use sampling because it was too expensive to measure the operating hours of all lighting circuits. Metering a sample of circuits provided an estimate of the true operating hours. To meet a 90/10 uncertainty criterion (confidence and precision) the sample size is determined such that, once the operating hours are estimated by sampling, the range of sample estimate (±10%) has to have a 90% chance of capturing the true hours of use. The conventional approach is to design sampling to achieve a 90% confidence level and ±10% precision. However, the M&V Plan needs to consider the limits created by the budget. For example, improving precision from ±20% to ±10% will increase sample size by four times, while improving it to ±2% will increase sample size by 100 times. (This is a result of the sample error being inversely proportional to.) Selecting the appropriate sampling criteria requires balancing accuracy requirements with M&V costs.
3. *Decide on the level of disaggregation.* Establish whether the confidence and precision level criteria should be applied to the measurement of all components, or to various sub-groups of components. For example, do you wish to have the 90/10 apply to a specific room type or to the building overall?
4. *Calculate the initial sample size.* Based on the information above, an initial estimate of the overall sample size can be determined using the following equation:



Where:

* no is the initial estimate of the required sample size.
* cv is the coefficient of variation, defined as the standard deviation of the readings divided by the mean. Until the actual mean and standard deviation of the population can be estimated from actual samples, 0.5 may be used as an initial estimate for cv.
* e is the desired level of precision.
* z is the standard normal distribution value for the desired confidence level. Use the following values for z: 1.96 for a 95%; 1.64 for 90%; 1.28 for 80%; and 0.67 for 50% confidence levels. If other confidence values are desired, consult with a z table from any statistics text book or online sources.

For example, for 90% confidence with 10% precision, and a cv of 0.5, the initial estimate of required sample size (no) is:



In some cases (e.g., metering of lighting hours or use), it may be desirable to initially conduct a small sample for the sole purpose of estimating a cv value to assist in planning the sampling program. Also, values from previous M&V work may be used as appropriate initial estimates of cv.

1. *Adjust the initial sample size estimate for small populations*. The necessary sample size can be reduced if the entire population being sampled is no more than 20 times the size of the sample. For the initial sample size example above, (no = 67), if the population (N) from which it is being sampled is only 200, the population is only three times the size of the sample. Therefore, the “Finite Population Adjustment” can be applied. This adjustment reduces the sample size (n) as follows:



Applying the finite population adjustment to the above example reduces the sample size (n) required to meet the 90%/±10% criterion to 50.

As the initial sample size (no) or the adjusted sample size (n) are determined using an assumed cv, it is critical to remember that the actual cv of the population being sampled may be different. Therefore, a different actual sample size may be needed to meet the precision criterion. If the actual cv turns out to be smaller than the initial assumption in step 4, the required sample size will be unnecessarily large to meet the precision goals. If the actual cv turns out to be larger than assumed, then the precision goal will not be met unless the sample size increases beyond the value computed above.

As sampling continues, the mean and standard deviation of the readings should be computed. The actual cv and required sample size should be recomputed. Recomputation may allow early curtailment of the sampling process. It may also lead to a requirement to conduct more sampling than originally planned. To maintain M&V costs within budget, it may be appropriate to establish a maximum sample size. If this maximum is actually reached after the above recomputations, the savings report(s) should note the actual precision achieved by the sampling.

## Determination of Metering Period

The determination of the metering period is often a very important consideration for any M&V project. Previous M&V work has shown that site-level metering should cover a full year whenever possible. Some research has indicated that six and nine month ranges have been sufficient, though this is highly dependent on the end-use and the building use. For instance, unconditioned warehouses may not need much more than a month of metering, while offices may require a few months of metering depending on the building use (i.e., if use is seasonal). For lighting, the metering period may not need to be long compared to heating or cooling measures.

The key determination is the likely variation throughout the year. Any metered load that exhibits weather dependence should ideally be metered over the relevant weather-dependent period. For instance, cooling loads should be metered over the summer and shoulder periods where temperatures exceed equipment set-points. Metering of educational facilities should capture both occupied and unoccupied periods. Industrial metering should cover sufficient variation in production levels to model the probable range of use. Daylight dependent loads should, at a minimum, span the six month period from solstice to solstice.

## Modeling

Modeling involves finding a mathematical relationship between dependent (y) and independent (explanatory) variables (x). Models attempt to *explain* the variations in Y using the explanatory variables provided by the modeler. For example, if y is energy use, then weather, size of the building, orientation, etc. may all be used to explain the observed energy use. This type of modeling is called multiple regression analysis.

Multiple regression equation takes the following form:

*y = b1x1 + b2x2 + ... + bnxn + c*

The bs are the regression coefficients, representing the amount the dependent variable y changes when the corresponding independent changes by 1 unit. The c is the constant, where the regression line intercepts the y axis, representing the amount the dependent y will be when all the independent variables are 0. For example, if the xs are weather variable, then c may be the amount of baseload that is not impacted by weather and the individual b values explain the amount of energy use that is attributable to weather (e.g., kWh per heating degree days). Associated with multiple regression is R2, which is the percentage of variation in the dependent variable explained collectively by all of the independent variables.

For example, let Y be energy use (usually in the form of energy use during a specific time period; one hour, one day, 30 days, etc.). Now let xi (i = 1, 2, 3, …, k) represents the k explanatory variables such as weather, production, occupancy, etc. bi (i = 0, 1, 2, … k) represents the coefficients derived for each independent variable. e represents the residual errors that remain unexplained after accounting for the impact of the various independent variables. The most common regression analysis finds the set of bi values that minimizes the sum of squared residual-error terms (these regression models are thus called least-squares models).

An example of the above model for a building’s energy use is:

*Monthly energy consumption = 342,000 + (63 x HDD) + (103 x CDD) + (222 x Occupancy)*

HDD and CDD are heating and cooling degree days, respectively. Occupancy is a measure of percentage occupancy in the building. In this model, 342,000 is an estimate of base load in kWh, 63 measures the change in consumption in kWh for one additional HDD, 103 measures the change in consumption in kWh for one additional CDD, and 222 measures the change in consumption in kWh per 1% change in occupancy.

### Coefficient of Determination (R2)

The first step in assessing the accuracy of a mode is to examine the Coefficient of Determination, R2, a measure of the extent to which variations in the dependent variable y are explained by the regression model. Mathematically, R2 is:



All statistical packages and spreadsheet regression-analysis tools compute the value of R2.

The range of possible values for R2 is 0.0 to 1.0. An R2 of 0.0 means none of the variation is explained by the model, therefore the model provides no guidance in understanding the variations in Y (i.e., the selected independent variable(s) give no explanation of the causes of the observed variations in Y). On the other hand, an R2 of 1.0 means the model explains 100% of the variations in Y, (i.e., the model predicts Y with total certainty, for any given set of values of the independent variable(s)). Neither of these limiting values of R2 is likely with real data.

In general, the greater the coefficient of determination, the better the model describes the relationship of the independent variables and the dependent variable. Though there is no universal standard for a minimum acceptable R2 value, as it is highly dependent on the context. Some portion of a model will always be unexplained and fit should be assessed with this in mind. For instance, R2 values tend to be higher in time-series analysis, where a single site is analyzed, as a site has less variation from time period to time period (within error) than multiple sites do between each other at any given time period (between error).

The R2 test should only be used as an initial check. Models should not be rejected or accepted solely on the basis of R2. Finally, a low R2 is an indication that some relevant variable(s) are not included, or that the functional form of the model (e.g., linear) is not appropriate. In this situation it would be logical to consider additional independent variables or a different functional form. However, variables should only be added in the case that they have theoretical merit, otherwise they could lead to bias in the estimates.

R2 value will always increase when you add additional explanatory variables, whether relevant or not. It never decreases. Consequently, a model with more explanatory variables may appear to have a better fit simply because it has more terms. In order to prevent modelers from falling in the trap of maximizing R2 at all costs, it is recommended that adjusted- R2 is used. The adjusted R2 is a modified version of R2 that has been adjusted for the number of predictors in the model. The adjusted R2 increases only if the new term improves the model more than would be expected by chance. It decreases when a predictor improves the model by less than expected by chance. The adjusted R2 can be negative, but it’s usually not.

### t-statistic

Since regression-model coefficients (bk) are statistical estimates of the true relationship between an individual X variable and Y, they are subject to variation. The accuracy of the estimate is measured by the standard error of the coefficient and the associated value of the t-statistic. A t-statistic is a statistical test to determine whether an estimate has statistical significance.

The standard error of each coefficient is computed by regression software. The following equation applies for the case of one independent variable.



For cases with more than one independent variable, the equation provides reasonable approximation when the independent variables are truly independent (i.e., not correlated with each other). Otherwise, the equation gets very complex and the M&V analyst is better off using a software package to compute the standard errors of the coefficients. The range within which the true value of the coefficient, b falls is found using:

*b ± t \* SEb*

The standard error of the coefficient, b, also leads to the calculation of the t-statistic. This test ultimately determines if the computed coefficient is statistically significant. The t-test is computed by all statistical software using the following equation:



A rule of thumb states that the absolute value of a t-statistic result of 2 or more implies that the estimated coefficient is significant, and therefore that a relationship does exist between Y and the particular X related to the coefficient. It can then be concluded that the estimated b is not zero. However, at a t-statistic of about 2, the precision in the value of the coefficient is about ±100%: not much of a vote of confidence in the value of b. To obtain a better precision of say ±10%, the t-statistic values must be around 20, or the standard error of b has to be no more than 0.05 of b itself.

All statistical software and spreadsheet provide p-values for the computed regression coefficient. Each t-test will have a corresponding p-value. P-value is a number between 0 and 1 and interpreted as:

* Small. Typically ≤ 0.05. This indicates strong evidence that the explanatory variable is significant (i.e., does have significant impact on the dependent variable).
* Large. Typically > 0.05. This is indicates weak evidence that the explanatory variable is significant.
* Close to 0.05 are considered to be marginal (could go either way).

To improve the t-statistic (and p-values) results considered the actions below:

* select independent variable(s) with the strongest relationship to energy;
* select independent variable(s) whose values span the widest possible range (if X does not vary at all in the regression model, b cannot be estimated and the t-statistic will be poor);
* gather and use more data points to develop the model; or
* select a different functional form for the model; for example, one which separately determines coefficient(s) for each season in a building that is significantly affected by seasonal weather changes.

## Modeling Errors

When using regression models, as described above, several types of errors may be introduced as listed below.

* The mathematical model may not include relevant explanatory variables (omitted variable bias).
* The model may include some variables that are irrelevant.
* The model may use inappropriate functional form.
* The model may be based on insufficient or unrepresentative data.

### Omission of Relevant Variables

In M&V, regression analysis is used to account for changes in energy use. Most complex energy using systems are affected by innumerable explanatory variables. Regression models cannot hope to include all independent variables. Even if it were possible, the model would be too complex to be useful and would require excessive data gathering activities. The practical approach is to include only explanatory variable(s) thought to significantly impact energy.

Omission of a relevant independent variable may be an important error. If a relevant independent variable is missing (e.g., HDD, production, occupancy), then the model will fail to account for a significant portion of the variation in energy. The deficient model will also attribute some of the variation that is due to the missing variable to the variable(s) that are included in the model. The effect will be a less accurate model.

Note that there are two possible consequences to omitted variable bias: less precision and possible bias. Precision is lost in the case where an omitted variable explains changes in energy consumption, but is not related to other explanatory variables included in the models. Bias is introduced in the case that an omitted variable is related to both the dependent and independent variable(s).

There are no obvious indications of this problem in the standard statistical tests (except maybe a low R2). Experience and knowledge of the engineering of the system whose performance is being measured is valuable in addressing this issue.

There may be cases where a relationship is known to exist with a variable recorded during the baseline period. However the variable is not included in the model due to lack of budget to continue to gather the data in the reporting period. Such omission of a relevant variable should be noted and justified in the M&V Plan.

### Inclusion of Irrelevant Variables

Sometimes models include irrelevant independent variable(s). If the irrelevant variable has no relationship (correlation) with the included relevant variables, then it will have minimal impact on the model. However, if the irrelevant variable is correlated with other relevant variables in the model, it may bias the coefficients of the relevant variables.

Use caution in adding more independent variables into a regression analysis just because they are available. To judge the relevance of independent variables requires both experience and intuition. However, the associated t-statistic is one way of confirming the relevance of particular independent variables included in a model. Experience in energy analysis for the type of facility involved in any M&V program is necessary to determine the relevance of independent variables.

### Functional Form

It is possible to model a relationship using the incorrect functional form. For example, a linear relationship might be incorrectly used in modeling an underlying physical relationship that is non-linear. For example, electricity consumption and ambient temperature tend to have a non-linear (often ‘U’ shaped) relationship with outdoor temperature over a one-year period in buildings that are both heated and cooled electrically. (Electricity use is high for both low and high ambient temperatures, while relatively low in mid seasons.) Modeling this non-linear relationship with a single linear model would introduce unnecessary error. Instead, separate linear models should be derived for each season.

It may also be appropriate to try higher order relationships, e.g., Y = f(X, X2, X3).

The modeler needs to assess different functional forms and select the most appropriate among them using evaluation measures. A useful tool in assessing model functional form is visualization. Often simply examining scatter plots can make the functional form of the model apparent.

### Data Shortage

Errors may also occur from insufficient data either in terms of quantity (i.e., too few data points) or time (e.g., using summer months in the model and trying to extrapolate to winter months). The data used in modeling should be representative of the range of operations of the facility. The time period covered by the model needs to include various possible seasons, types of use, etc. This may call for either extension of the time periods used or increasing sample sizes.

### Evaluating Regression Models

In order to evaluate how well a particular regression model explains the relationship between energy use and independent variable(s), three tests may be performed as described below.

#### Prediction Errors

When a model is used to predict an energy value (Y) for given independent variable(s), the accuracy of the prediction is measured by the standard error of the estimate (SE). This is also known by a couple of other names, standard error of the regression and root-mean-squared error adjusted for degrees of freedom. In regression modeling, the best single error statistic to look at is the standard error of the regression, which is directly related to the unexplainable variations in the dependent variable (Y). This what your statistics software is trying to minimize when estimating coefficients.

Once the value(s) of the explanatory variable(s) are plugged into the regression model to estimate an energy value (), an approximation of the range of possible values for  can be computed using:



Where:

* is the predicted value of energy (Y) from the regression model
* t is the value obtained from the t-tables (see Table 1)
* is the standard error of the estimate (prediction). It is computed as:



In this equation, k is the number of explanatory variables in the regression equation. Again, this often also referred to as the root-mean squared error (RMSE). The RMSE is a quadratic equation that measures the average magnitude of the error. The difference between the estimated and corresponding observed values each are squared and then averaged over the sample. The square root of the average is taken. RMSE gives a relatively high weight to large errors. Dividing the RMSE by the average energy use produces the coefficient of variation of RMSE, or the CV(RMSE), as shown here:



All these measures may be used in evaluating the calibration of simulation models in Option D.

### Overfitting

Overfitting occurs when an overly complex model is estimated to the data. This can lead parameters representing not just the signal but also the noise in the data. This will lead to models that do a poor job of predicting results in other contexts (i.e., outside of the data range in the model). While M&V models are not typically used for forecasting, they are used to generate estimates of typical savings expected over time. Therefore it is important to estimate that explain enough of the variation in energy usage without overfitting.

A good way to protect against overfitting is by using cross validation techniques. In cross-validation, a portion of the data collected is set aside and not used to estimate the model. Once the model specification is chosen, it is tested against the remaining data. R2 and prediction errors can be computed for the test data to give a better estimate of the likely out-of-sample error

## Combining Components of Uncertainty

If the reported savings is the sum or difference of several independently determined components (C) (i.e., ), then the standard error of the reported savings can be estimated by:

SE(*Savings*) = 

For example, if savings are computed using the equation:

*Savings = (Baseline Energy – Reporting-Period Energy) ± Adjustments*

the standard error of the difference (savings) is computed as:

*SE*(*Savings*) = 

If the reported savings estimate is a product of several independently determined components (i.e., ), then the relative standard error of the savings is given approximately by:

 (20)

A good example of this situation is the determination of lighting savings as:

*Savings = Δ Watts x Hours*

If the M&V Plan requires measurement of hours of use, then “Hours” will be estimated and will have a standard error. If the M&V Plan also includes measurement of the change in wattage, then ΔWatts will also be a value with a standard error. Relative standard error of savings will be computed using the formula above as follows:



The components must be independent to use the methods above for combining uncertainties. Independence means that whatever random errors affect one of the components are unrelated to the errors affecting other components.

The remainder of this document is divided into three main sections:

* Section I, covering Options A and B and is further divided into four sections as illustrated in figure xx below. The division into these particular sections is based on the type of use (variable or constant) and ability to meter all units. Each of the four subsections describes, through an example, the assessment and treatment of uncertainty.
* Section II, covering Option C
* Section II, covering Option D

We believe these sections and associated examples cover the great majority of M&V scenarios.

Figure . Section I (Options A and B)

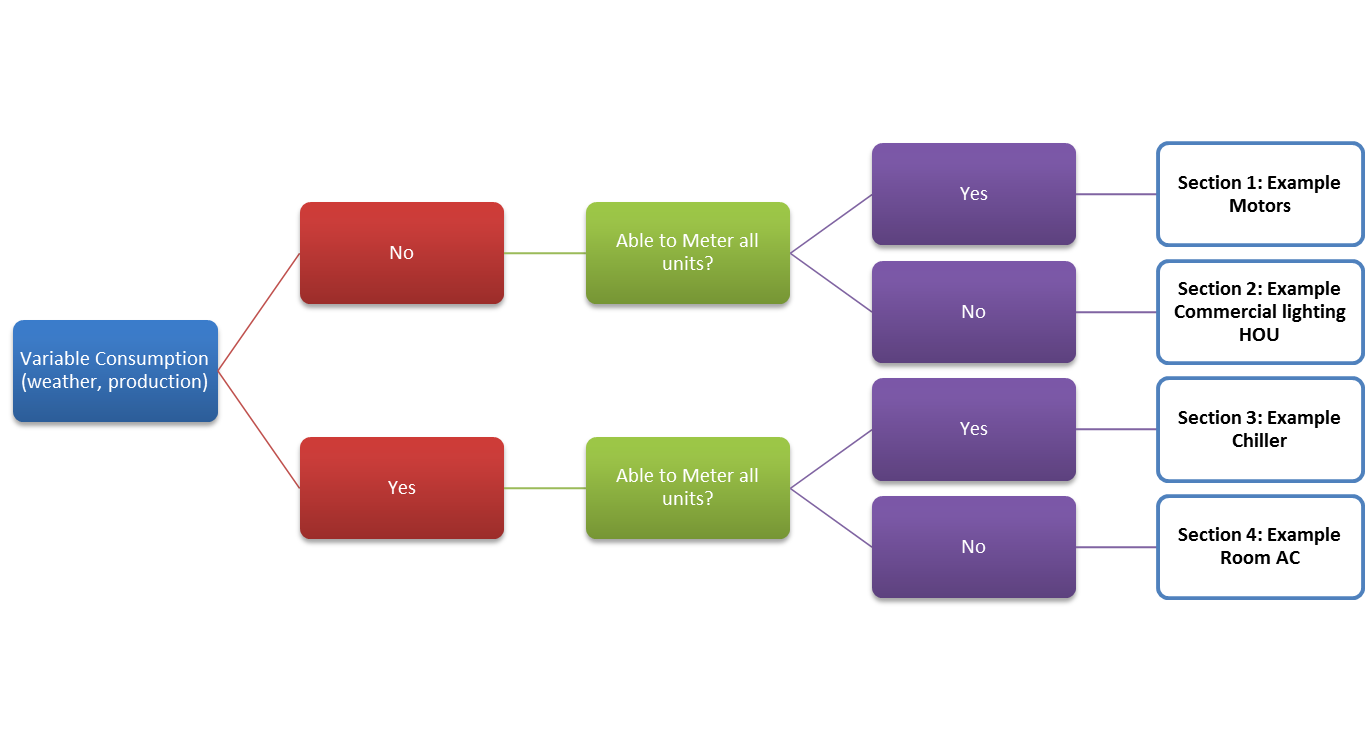


Figure . Section II (Option C)

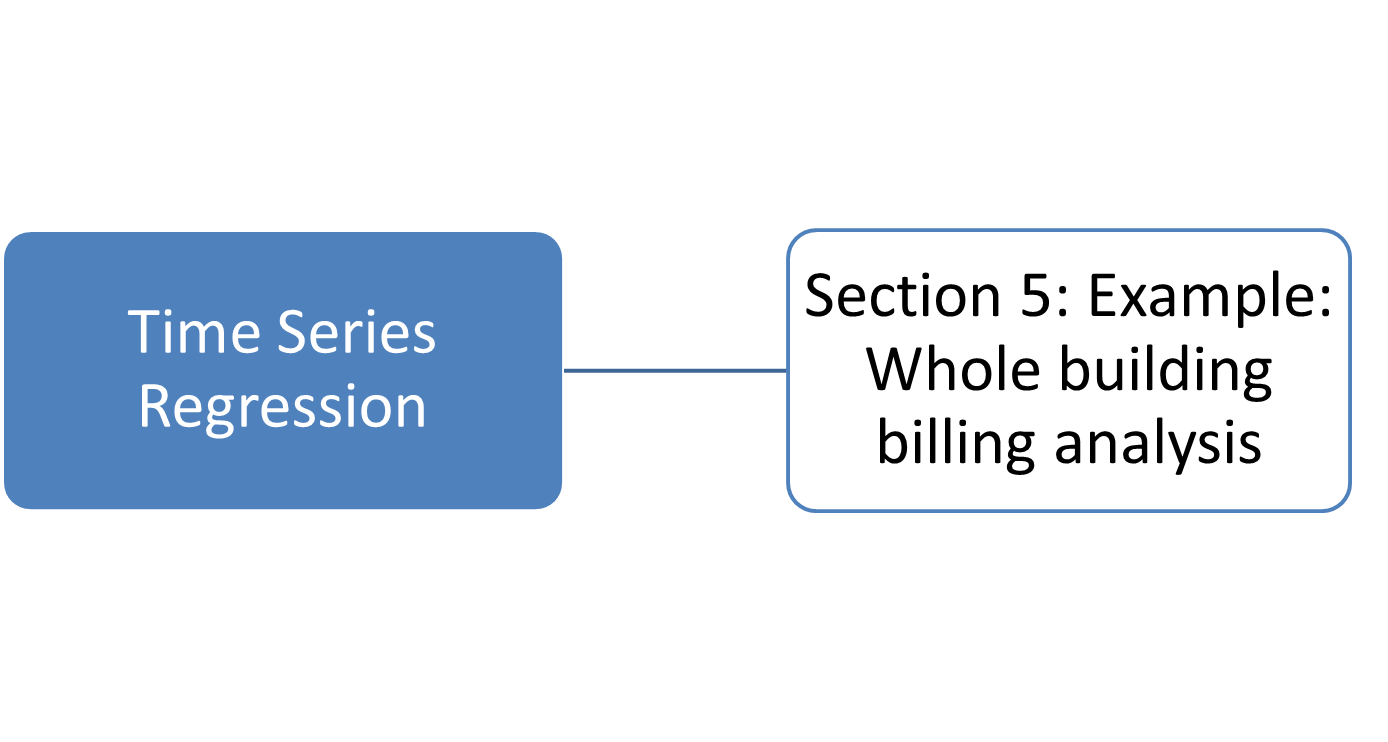


Figure . Section II (Option D)

