8 Measurement and Verification Using Billing Analysis and Regression Models

8.1 Overview

Billing analysis involves the use of regression models with historical utility billing data (kW and kWh) to calculate annual demand and energy savings. In general, billing analysis is used with complex equipment retrofits and controls projects. Examples of the types of projects where billing analysis may be employed include the installation of an energy management control system (EMCS), and a comprehensive building retrofit involving multiple types of energy efficiency measures (EEMs).

Billing analysis provides retrofit performance verification for projects where whole-facility baseline and post-installation data are available. Billing analysis usually involves collection of historical whole-facility baseline energy use data and a continuous measurement of the whole-facility energy use after measure installation. Baseline and periodic inspections of the equipment may also be warranted. Energy consumption is calculated by developing statistically representative models (multivariate regression models) of historical whole-facility energy consumption (kWh).

The M&V method described here is based, in part, on Option C of the 2001 International Performance Measurement and Verification Protocol (IPMVP). Valuable insights on utility bill analysis can be found in the IPMVP.

8.2 Baseline and Post-Retrofit Data Collection

Collecting and validating data, as well as ensuring alignment of data start and end dates are important elements of billing analysis. Data types and some data analysis protocols are discussed below.

8.2.1 Data Types

As input to the multivariate regression models, billing data provide the basis for calibrating models and post-installation energy use. Site data provide a means for controlling changes in energy use not associated with measure installation. These data elements are discussed below.

- **Monthly Energy Billing Data.** There are typically two types of monthly energy billing data; total energy usage for the month, or energy usage aggregated by time-of-use periods. While either type of data can be used with a regression model, time-of-use is preferable as it provides more insight into usage patterns.
- **Interval Demand Billing Data.** This type of billing data records the average demand for a given interval (e.g., 15 minutes) associated with the billing period.
- **Site Data.** Site data provide the information necessary to account for either changes in or usage of energy consumption that is not associated with the retrofit equipment. Typical site data that can be incorporated in regression models include weather parameters, occupancy, facility square footage and operating hours. These data are typically used to help define the independent variables that explain energy

consumption or change associated with equipment other than the equipment installed as part of an EEM.

8.2.2 Data Analysis Protocols

The following are some of the required data analysis protocols:

- **Baseline Energy Consumption.** This regression analysis requires at least 12 months' worth of data prior to the date of installation. However, if energy consumption is sensitive to weather, or other highly variable factors, then at least 24 months worth of data are required.
- **Post-installation Energy Consumption.** This regression analysis requires at least nine months, and preferably twelve months of data after the date of installation to determine impacts for the first year.
- **Outliers.** Outliers are data beyond the expected range of values (e.g., a data point more than two standard deviations away from the average of the data). However, the elimination of outliers should be explained. It is not sufficient to eliminate a data point because it is beyond the expected range of values. If there is reason to believe that the data point is abnormal because of specific mitigating factors, then it can be eliminated from the analysis. Nevertheless, if a reason for the unexpected data point cannot be found, it should be included in the analysis. Outliers should be defined based on "common sense" as well as common statistical practice. Outliers can be defined in terms of consumption changes and actual consumption levels.

8.3 Calculation of Energy Savings: Multivariate Regression Method

Multivariate regression is an effective technique that controls for non-retrofit-related factors that affect energy consumption. If the site data (all relevant explanatory variables, such as weather, occupancy, and operating schedules) are available and/or collected, the technique should result in more accurate and reliable savings estimates than a simple comparison of pre- and post-installation energy consumption.

The use of the multivariate regression approach is dependent on and limited by the availability of site and billing data. The decision to use a regression analysis technique should be based on the availability of this information. Thus, on a project-specific basis, it is critical to investigate the EEM dependent and independent variables that have direct relationships to energy use. Data need to be collected for these variables in a suitable format over a significant period of time.

Separate models may be proposed that define pre-installation energy use and post-installation energy use with savings equal to the difference between the two equations. It is assumed, however, that a single "savings" model will be simpler and generate more reliable estimates since it is also based on more data points.

8.3.1 Overview of the Regression Approach

Regression models should be developed that describe pre-installation and post-installation energy use for the affected site (or sites), taking into account all explanatory variables.

For projects with time-of-use utility billing data, the regression models should yield savings by hour or critical time-of-use period. For projects with only monthly consumption data, the models should be used to predict monthly savings.

8.3.2 Standard Equation for Regression Analysis

In the regression analysis, utility billing data (monthly or hourly) on a project-specific basis are used to develop the models for comparing the pre-installation energy use to post-installation energy use. After adjusting for non-retrofit-related factors in the models, the models' energy use difference is defined as the gross performance impact of the EEMs.

The regression equations should be specified so as to yield as much information as possible about savings impacts. For example, with hourly data, it should be possible to estimate the savings impacts by time of day, day of week, month, and year. With only monthly data, however, it is only possible to determine the effects by month or year. Data with a frequency lower than monthly should not be used under any circumstances.

8.3.3 Independent Variables

Independent variables that affect energy consumption should be specified for use in the regression analysis. These variables can include weather, occupancy patterns, and operating schedules.

If the multivariate regression models discussed above incorporate weather in the form of heating degree-days (HDD) and/or cooling degree-days (CDD), the following issues must be considered:

- The use of the building "temperature balance point" for defining degree-days versus an arbitrary degree-day temperature base.
- The relationship between temperature and energy use that tends to vary depending upon the time of year. For example, a temperature of 55°F in January has a different implication for energy usage than the same temperature in August. Thus, seasonality should be addressed in the model.

8.3.4 Testing Statistical Validity of Models

The statistical validity of the final regression model should be tested by the Sponsor and CenterPoint Energy or its contractor and should demonstrate the following:

- The model makes intuitive sense; e.g., the independent variables are reasonable, and the coefficients have the expected sign (positive or negative) and are within an expected range (magnitude).
- The modeled data are representative of the population.
- The form of the model conforms to standard statistical practice.
- The number of coefficients is appropriate for the number of observations (approximately no more than one explanatory variable for every five data observations).
- The T-statistic for all key parameters in the model is at least 2 (95% confidence that the coefficient is not zero).
- The model is tested for possible statistical problems and, if present, appropriate statistical techniques are used to correct for them.

- All data input to the model are thoroughly documented, and model limits (range of independent variables for which the model is valid) are specified.

8.3.5 Compliance with Energy Standards

When using billing analysis methods, the baseline should comply with minimum state and federal energy standards with respect to the following:

- Baseline equipment/systems should not include devices (e.g., lamps and ballasts) that are not allowed to be installed under current regulations.
- Baseline equipment should meet *prescriptive* efficiency standards requirements for affected equipment.
- Surveys and analysis correction methods (potentially outside of the model) should be documented in a project-specific M&V plan.
- The baseline *does not* have to comply with *performance compliance* methods that require the facility to meet an energy budget.

8.3.6 Detailed Calculation Issues

The details of the savings calculations are dependent on such issues as:

- The use of hourly versus monthly utility meter billing data
- The format of the data (e.g., corresponding to same time interval as the billing data) and availability of *all* relevant data for explanatory variables
- The amount of available energy consumption data
- The use of actual or typical data to calculate savings
- Compliance with energy standards when calculating baseline energy use. Energy savings should be calculated with the incorporation of minimum state and federal energy efficiency standards or codes into the determination of baseline energy use.

8.4 Project Specific M&V Issues

When billing analysis methods are used, the project specific M&V plan should address, in addition to other topics generic to all M&V methods, the following:

- How billing data covering an adequate period of time should be used to calculate savings in the performance year?
- How the baseline will be adjusted in order to have the baseline meet minimum energy standards?