Oracle Utilities Opower Evaluation, Measurement, and Verification (EM&V) Data Extract

Randomized Controlled Trials

E84793-01

Specification Version Number: v1-6-0

Last Update: January 10, 2019

ORACLE®

Oracle Utilities Opower Evaluation, Measurement, and Verification (EM&V) Data Extract

E84793-01

Documentation Accessibility

For information about Oracle's commitment to accessibility, visit the Oracle Accessibility Program website at http://www.oracle.com/pls/topic/lookup?ctx=acc&id=docacc.

Access to Oracle Support

Oracle customers that have purchased support have access to electronic support through My Oracle Support. For information, visit http://www.oracle.com/pls/topic/lookup?ctx=acc&id=info or visit http://www.oracle.com/pls/topic/lookup?ctx=acc&id=trs if you are hearing impaired.

Version History

The following table lists the updates to this document by date.

Update	Specification Version Number	Date
Edit sections and headings to simplify content and clarify instructions.	v1-6-0	01/10/2019
Updated file retention section to reflect regional differences.	v1-5-0	05/11/2018
Added information about Oracle file retention policy.	v1-4-0	03/12/2018
Updated data file descriptions to reflect changes to the schema.	v1-3-0	08/09/2017
Updated document organization and styling.	v1-2-0	04/06/2017
Updated branding.	v1-1-0	02/17/2017
Added baseline specification version number to the document. Added legal notices. Updated document branding.	v1-0-0	01/11/2017
Initial publication.	Not applicable.	06/30/2014

Contents

Introduction

Oracle Utilities Opower Evaluation, Measurement & Verification (EM&V) data extract files for randomized controlled trials (RCTs) contain the customer and usage data necessary for third-party evaluation and verification of energy savings. Oracle Utilities provides EM&V data extracts for RCTs upon utility client request. This documentation outlines the format and contents of the EM&V data extract, explains the data preparation procedures, and describes how Oracle Utilities uses this prepared data to estimate savings.

Data Extract Specification

The provided EM&V data file contains raw customer and billing data without any data preparation steps taken. Data columns are defined in the table below.

Column Name	Description
opower_customer_id	Oracle Utilities unique customer identifier.
measurement_name	The name of the savings measurement.
study_name	A single measurement may consist of multiple randomized controlled trials (RCT) or groups of customers. Savings for each RCT or group of customers must be calculated independently and then aggregated into a final, single result.
utility_code	Oracle Utilities client abbreviation.
treatment	Artificial variable indicating the treatment status of a customer: • 0 = Control • 1 = Treatment
rct_start_date	Date marking the start of the RCT program. This is defined as the beginning of the month in which reports are first generated for a cohort. For rolling enrollment waves, a cohort consists of all customers randomized into the wave at the same time.
customer_rct_end_date	Date that the customer was removed from the RCT program

Column Name	Description
	(for example, due to being inactive). This value will be NULL if the customer is still in the RCT program.
acct_active_date	Customer account active date.
acct_inactive_date	Customer account inactive date. This value will be NULL if the account is still active.
bill_date	Date meter was read.
usage_units	Meter usage units: • "kwh" • "therm"
usage_value	Amount of energy used over bill duration.
bill_duration_days	Bill duration in days.
is_estimate	Variable indicating whether the meter read is an estimate or not. • 0 = Actual read • 1 = Estimated read **Note**: Actual reads following estimated reads contain usage for the bill period and additional usage for truing-up any error in the estimate. See Estimated Read True-Up for more information.

Data Preparation Procedures

This section describes the data preparation procedures used by Oracle Utilities to transform the provided raw dataset into a panel dataset which can be entered into a regression model.

2

Estimated Read True-Up

Some utilities estimate usage for a billing period to save on operational costs. A subsequent read provides the actual usage for the billing period. To increase measurement accuracy, Oracle Utilities trues-up original usage estimates by first finding the estimate correction, which is the difference between the actual read and estimated read. Then, the estimate correction is added to the original estimate to get actual usage for the billing period. See the image below for more information.
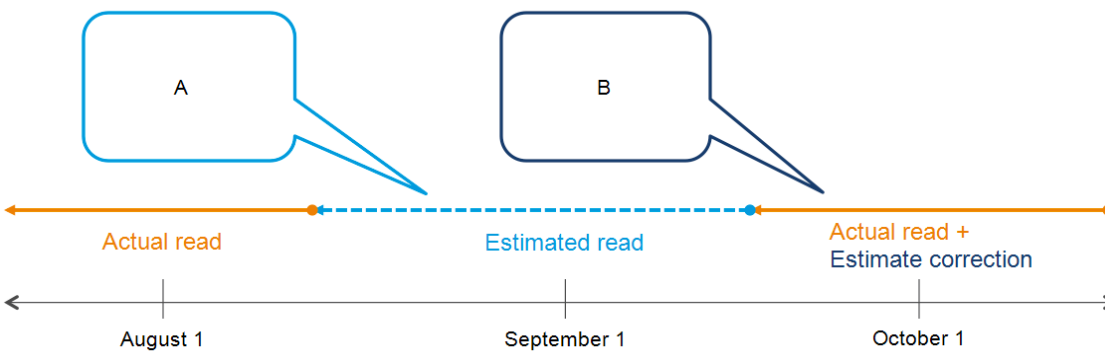
A. The bill for August 15 through September 15 is based on estimated usage.
B. The actual read on October 15 consists of actual usage over the period of September 15 through October 15 plus an estimate correction.

In the image above, truing-up is the function of adding the usage over periods A and B to obtain actual usage from August 15 through October 15.

Because Oracle Utilities does not know if a customer's first non-estimated read in the raw billing dataset includes an estimate correction to true-up prior estimates, the first non-estimated bill for each customer and all estimated reads prior to that bill are not included in the prepared dataset.

Billing Calendarization

Calendarization is the process of pro-rating billing data into calendar months. Calendarization can smooth out billing data with read durations longer than one month or billing data with a significant percentage of estimated reads.

For example, a 30-day bill dated August 15 includes usage that occurred in the second half of July, but does not include usage from the second half of August. The image below demonstrates the span of this billing period.

Billing period
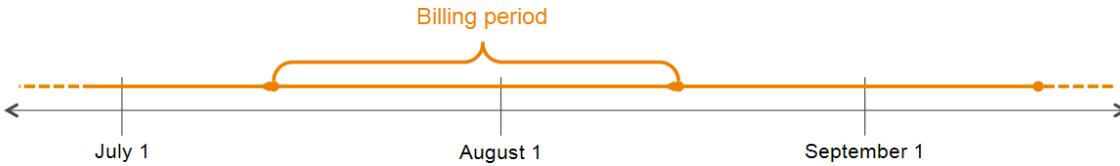
Previously, all usage in this period would be attributed to August for savings calculations, as shown in the image below.

Before calendarization
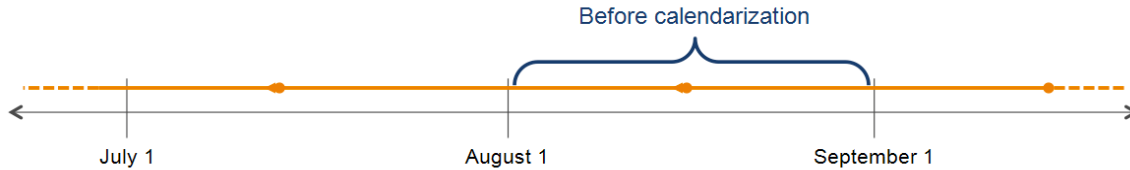
Calendarization spreads usage into the calendar month in which it took place. Using calendarization, 15 out of the 30 days of usage are attributed to July and the remaining 15 days of usage are attributed to August. The image below shows how the process of calendarization affects the calendar month to which usage data is attributed.

After calendarization
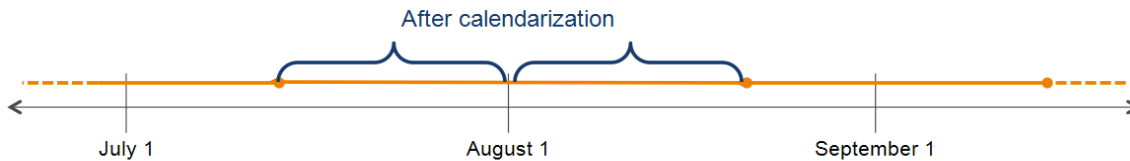
Trim Billing

After calendarization, Oracle Utilities trims usage data to exclude outliers and remove inapplicable data. The following criteria is used to identify outliers:

- Usage occurring after the customer move out date.
- Billing data with duration less than 1 day or more than 31 days (to trim months with overlapping meter reads).
- Usage less than -300 kWh per day, greater than 300 kWh per day, less than -50 therms per day, or greater than 50 therms per day.
- The most recent partial month. For example, if analysis is run in December, only usage through November is included in analysis.
- Usage occurring more than 12 months prior to the program start date.

Average Pre-Treatment Usage Variables

The regression model uses billing data in the pre-treatment period to create three regression coefficients.

- **Average usage per day**: The average usage for a customer in the pre-treatment period (any meter reads that end prior to the `treatment_start_date`).
- **Average usage per day in summer**: The average usage for a customer in the pre-treatment period over the summer months, defined as June through September.
- **Average usage per day in winter**: The average usage for a customer in the pre-treatment period over the winter months, defined as December through March.

Estimate Savings

This section describes how to estimate savings from the panel dataset created from the provided raw data. It includes the model specification used by Oracle Utilities for regression analysis to calculate savings per day. It also includes instructions for estimating the total savings across the post-treatment period and the savings for each month.

Model Specification

The model specification is used in conjunction with data from the prepared dataset for regression analysis. Regression analysis yields β, the savings per day per household coefficient. In the model specification below, i denotes the ith household and t denotes the tth month of the post-treatment period.

$$daily\_usage_{it} \: = \: \alpha + \beta treatment_i \: + \: _{\gamma t} Y_{oi} \: + \: mm_t \: + \: \varepsilon_{it}$$

where

- $daily\_usage_{it}$ is the average daily usage for meter read t for household i in the post-treatment period.
- $treatment_i$ is an indicator that household i is assigned to the treatment group.
- Y_{oi} is a vector of three baseline usage control variables:
  - $avg\_preusage_i$ is the average daily usage across household i's available pre-treatment meter reads. If missing, this value is imputed with the average pre-usage of all customers in the population.

- $avg\_preusage\_winter_i$ is the average daily usage over the months of December through March across household i's available pre-treatment meter reads. If missing, this value is imputed with household i's value for $avgpreusage_i$.

- $avg\_preusage\_summer_i$ is the average daily usage over the months of June through September across household i's available pre-treatment meter reads. If missing, this value is imputed with household i's value for $avgpreusage_i$.

- mm_t is a set of indicators for each month-year in the post-treatment period.

- γ^t coefficients represent the relationship of the *avg_preusage* control variables to post-treatment usage. γ^t coefficients can vary by month-year.

- ε_{it} is an error term.

The regression is weighted by *bill_duration_days* in order to reflect that billing records have been averaged over a varying number of days to calculate *daily_usage*.

Implementing the Model Specification

The following can be used to implement the model specification in R:

Cumulative model:

```
daily_usage ~ treatment + mm*(avg_preusage + avg_preusage_summer
+ avg_preusage_winter)
```

Monthly model:

```
daily_usage ~ treatment:mm + mm*(avg_preusage + avg_preusage_
summer + avg_preusage_winter)
```

The following can be used to implement the model specification as a Stata command:

Cumulative model:

```
reg daily_usage treatment i.mm##c.avg_preusage ///

            i.mm##c.avg_preusage_summer ///

            i.mm##c.avg_preusage_winter ///

            [pw = bill_duration_days] ///

            if post == 1 & trimmed == 0 ///

            , vce(cluster opower_customer_id)
```

Monthly model:

6

```
reg daily_usage c.treatment#i.mm i.mm##c.avg_preusage ///

            i.mm##c.avg_preusage_summer ///

            i.mm##c.avg_preusage_winter ///

            [pw = bill_duration_days] ///

            if post == 1 & trimmed == 0 ///

            , vce(cluster opower_customer_id)
```

Calculating Cumulative and Monthly Savings

The following procedure describes how to use the prepared panel data and model specification to calculate cumulative and monthly savings in regression analysis.

To calculate cumulative and monthly savings:

1. Include in the regression all rows after the treatment start (billing data prior to treatment start should be used only to create regression coefficients).
2. Estimate the savings per day coefficient (β) for each month using the model specification.
3. Calculate the number of active customer days for each month by using the customer account active and inactive dates and the customer RCT start and end dates.
4. Estimate monthly savings by multiplying savings per day (β) for each month by the number of active days in each month. Estimate cumulative savings over the whole post-treatment period by multiplying savings per day (β) across all active days in the post-treatment period.

Data Extract File Transfer

Oracle Utilities uses a specific format, naming convention, and file transfer method to send EM&V data extracts to utilities.

File Format: Each EM&V data extract for an RCT is provided to third-party evaluators as a tab-delimited text file. The file contains only raw customer and billing data, leaving all data preparation decisions to the third-party evaluator.

File Name Convention: Files transferred from Oracle Utilities follow a standard file name convention.

<measurement name>_raw.tsv

File Transfer: Files are delivered to the Oracle UtilitiesOpower Secure File Transfer Protocol (SFTP) upon client request.

File Retention Policy

According to the retention policy described in the *Oracle Cloud Hosting and Delivery Policies* document (which can be found online at <u>Oracle Cloud Services</u>), all incoming and outgoing files added to the SFTP server will be retained for a maximum of 60 days. After this time period, they will be automatically deleted. Copies of files added to the SFTP server are stored in a file archive until the termination of the contract. For utilities in the European Union, files are retained for a maximum of 30 days in both the SFTP server and the file archive. After this period the files are deleted.