CASCADE NATURAL GAS LOAD STUDY ANALYSIS

DESIGN DOCUMENT

FEBRUARY 2021

This document describes the data sources and analysis techniques that Cascade anticipates using for its load study.

TABLE OF CONTENTS

OVERVIEW

The purpose of this document is to discuss the data gathering, estimation techniques and supporting documentation to complete a load study for Cascade Natural Gas Corporation (“Cascade” or “Company”). The objective of the load study is to quantify and assign peak responsibility by comparing individual group demands to system demands. Once Cascade has performed its load study for all customer groups, Cascade will be able assign service costs for any class or customer based on their contribution to the system peak.

DATA AND DATA SOURCES

The data inputs for Cascade’s load study include customer usage data, customer count, and weather data.

USAGE DATA

Cascade will analyze customer usage data in its load study analysis to determine the appropriate allocation for each rate schedules usage on peak day.

Cascade’s customer usage data for the load study is being captured at the per customer per day level and will be gathered through Cascade’s online database called ThoughtSpot. Since the usage data is captured at that level, the rate schedule will be easily identifiable for this analysis.

CUSTOMER COUNT DATA

Cascade will analyze customer count data to determine the representative sample for the load analysis.

For each set of customer usage data, there will be one customer behind that data. The remaining number of customers will be determined by utilizing Cascade’s customer count file maintained by Cascade’s accounting department. If forecasted customer counts are needed, Cascade will utilize the forecasted customer counts from the Integrated Resource Plan (IRP).

WEATHER DATA

Cascade will utilize National Oceanic and Atmospheric Administration (NOAA) to gather daily weather data information. This data is gathered at seven weather locations: four in Washington and three in Oregon. The four in Washington are Bellingham, Bremerton, Yakima, and Walla Walla. The three in Oregon are Baker City, Redmond, and Pendleton. Normal wind is defined as the average daily wind speed and normal weather is defined as the average daily temperatures of the most recent 30 years of historical data which results in the average annual temperatures as well.

The company uses a heating degree day (HDD) as the unit of measure for temperature. HDD is calculated by taking the average temperature from a day and subtracting it from a reference temperature. If the reference temperature less HDD is negative, then the company gives that day a 0 for HDD. The company uses 60°F as the reference temperature. For example, a 50°F day will result in 10 HDDs (60-50).

SAMPLE SIZE

Cascade plans to utilize Confidence Intervals and Relative Precision to determine the appropriate sample size. PURPA suggests that electrical utilities should maintain a minimum 90 percent confidence interval with a 10 percent accuracy at time of class and system peaks.[1] Cascade believes these values can be utilized for natural gas as well. To determine the initial sample size, Cascade plans to utilize the Cochran

[1] Load Research Manual, 3rd edition, 2017, page 61: "…The desired accuracy should be determined for the study. The desired accuracy is usually expressed as a relative precision with a given confidence level. A design accuracy of +/- 10% at the 90% confidence level (often abbreviated "90/10") at the system and class peak time was specified in the Public Utility Regulatory Policy Act (PURPA) for all major rate classes (see Chapter 1 for additional background). This means that the sample will be designed to achieve a 90% confidence interval that is no more than 10% above and below the estimate of the mean.1 Although these federal standards were withdrawn in 1992, the PURPA specification remains load research common practice, particularly for samples that will be used to support rate cases or other regulatory requirements…"

formula once the entire population has been stratified. Stratification is the process of dividing the population into subpopulations which are generally more homogeneous. For example, the customers can be broken out by rate schedule and weather zone. Once the sample size is chosen, Cascade will use the actual usage data to verify the sample size is large enough. Once the Company determines the sample size is appropriate, Cascade will use estimation techniques to analyze the data.

ESTIMATION TECHNIQUES

Cascade has several techniques to choose from when analyzing a sample set of data. Each technique provides a different analysis with the same end goal result to determine the peak allocator within Cascade's cost of service study. However, each technique would provide varying results so deciding which technique to use will be important. Cascade will be holding stakeholder meetings to discuss the data collected as well as the different estimation techniques in mid-2021. The techniques are described below:

REGRESSION

Regression models are most commonly used to describe a relationship between explanatory variables and response variables.

Model:

$$\frac{Therms}{C_{Class}} = \alpha_0 + \alpha_1 HDD^D + \alpha_2 I_w + \alpha_3 WIND^D + Fourier(k) + ARIMA(p,d,q)$$

Model Notes:

- $\frac{Therms}{C_{Class}} = Usage\ by\ customer\ class.$
- $HDD^D = Daily\ Heating\ Degree\ Days\ from\ Weather\ Location.$
- $w = Weekend\ Indicator$
- $WIND^D = Daily\ Wind\ Speed\ Average\ from\ Weather\ Location.$
- $ARIMA\epsilon(p,d,q) = Indicates\ that\ the\ model\ has\ p\ autoregressive\ terms, d\ difference\ terms, and\ q\ moving\ average\ terms.$
- $Fourier\ Term = Indicates\ k\ number\ of\ seasonalities\ within\ dataset$

Cascade begins by running a simple linear model regressing on HDD's and wind by month. Then, the residuals are analyzed using the Durbin-Watson test to check for autocorrelation. If found, the model then adds an ARIMA term and a Fourier term. In order to run Fourier terms alongside ARIMA terms, the ARIMA term must be forced to not difference the data. Cascade pursues this process for each rate schedule and weather zone. As with the customer forecast, Cascade uses AIC, among other statistics, in determining which model to use. Below are pros and cons of this methodology:

Pro: The methodology gives the best estimation of usage while utilizing explanatory variables. Con: The methodology is more complex and time consuming than previous methodologies, requiring advanced analytical understanding of regression concepts. This model also requires utilization of a statistical software package. Cascade uses the R programming language widely used among statisticians and data miners for developing statistical software and data analysis.

MEAN-PER-UNIT

The mean-per-unit technique calculates an estimated demand per customer and expands this estimate to the target population by multiplying by the number of customers in the population.

Model:

$$\bar{y} = \sum_{i=1}^{n} \frac{y_i}{n}$$

Model notes:

- $\bar{y} = Sample\ mean\ of\ y\ variable.$
- $y_i = usage\ for\ sample\ i$
- $n = number\ in\ sample\ size$

Once this analysis is complete, the Company would apply $\bar{y}$ to the entire population (customers inside and outside of the sample size) to calculate the total usage for the customer class. Below are pros and cons of this methodology:

Pro: The methodology favors efficiency and speed and can be easily understood.

Con: The methodology does not account for the fact that usage will vary with weather. For example, this methodology could be useful if the sample size is limited to data within a narrow range of HDDs. This methodology would require a very large sample size so the model can break the data down into different ranges of HDDs. Again, as an example, usage data could be grouped together for similar weather events like 45-50 HDDs. This would require a few days of 45-50 HDDs to be recorded for the Company to analyze usage with 45-50 HDDs.

RATIO

Ratio estimation is a technique that can take advantage of the correlation of the x and y variables to obtain increased precision.

Model:

$$r = \frac{\sum_{i=1}^{n} y_i}{\sum_{i=1}^{n} x_i}$$

Model notes:

- $r = ratio\ of\ y\ and\ x$
- $y_i = usage\ for\ sample\ i$
- $x_i = HDD\ for\ sample\ i$

Once this analysis is complete, the Company would apply r to the entire population (customers inside and outside of the sample size) to calculate the total usage by HDD for each customer class. Below are pros and cons of this methodology:

Pro: The methodology favors efficiency and speed and can be easily understood easily understood. Circumstances lead this method to be used when the x and y are highly linearly correlated; thus, are used when known survey sample population variables to improve the known population total for the same variable in the frame.

Con: The methodology does not capture trends or estimates as well as a regression model would. You could only use one explanatory variables for the denominator in this model.

FINAL PRODUCT

Once a model is decided on, the analyst utilizes the chosen model for a customer class and weather area to calculate the total usage for the remaining population. For example, if Cascade runs a regression on a sample of data from residential customers in the Bellingham weather area, Cascade can use that regression for the remaining population of residential customers in the Bellingham weather area. This method continues for each rate schedule and weather area to account for every customer. Once complete, Cascade will have peak day usage estimates for all customers in each rate schedule. The allocation between rate schedules can then be calculated for the cost of service study.